**UNIVERSIDAD AUTÓNOMA DE MADRID**

**FACULTAD DE CIENCIAS**

**DEPARTAMENTO DE BIOLOGÍA MOLECULAR**

# CHARACTERIZATION OF THE DIDO3-SFPQ AXIS IN ALTERNATIVE SPLICING

**TESIS DOCTORAL**

**Carmen Mora Gallardo**

Madrid, 2019

**UNIVERSIDAD AUTÓNOMA DE MADRID**

**FACULTAD DE CIENCIAS**

**DEPARTAMENTO DE BIOLOGÍA MOLECULAR**

# CHARACTERIZATION OF THE DIDO3-SFPQ AXIS IN ALTERNATIVE SPLICING

**Carmen Mora Gallardo**

Licenciada en Biotecnología

Para optar al grado de Doctor en Biociencias Moleculares con Mención Internacional por la Universidad Autónoma de Madrid

| **Director:** | **Co-director:** | **Tutor:** |
|---|---|---|
| **Dr. Karel H.M. van Wely** | **Prof. Carlos Martínez Alonso** | **Dr. Alberto Martínez Serrano** |

**CENTRO NACIONAL DE BIOTECNOLOGÍA – CSIC**

MADRID, 2019

**El trabajo presentado en esta memoria ha sido realizado en el Departamento de Inmunología y Oncología del Centro Nacional de Biotecnología bajo la dirección del Dr. Karel H.M. van Wely y del Prof. Carlos Martínez Alonso.**

**Carmen Mora Gallardo ha sido beneficiaria de una beca FPI (FPI BES-2014-068580) concedida por el Ministerio de Economía, Industria y Competitividad (actual Ministerio de Ciencia, Innovación y Universidades)**

A mis padres, Juan Manuel y Lola,

porque siempre siempre están ahí.

**A theory is something nobody believes,**
**except the person who made it.**
**An experiment is something everybody believes,**
**except the person who made it.**

Albert Einstein

# INDEX

# INDEX

# ABSTRACT

# ABSTRACT

Alternative splicing is a fundamental mechanism that allows the generation of multiple isoforms of the same pre-mRNA. Although splicing is a necessary process in all eukaryotes from yeast to mammals, alternative splicing is more prevalent in higher than lower eukaryotes, and especially in vertebrates. Nearly 95% of mammalian genes undergo alternative splicing, so mis-regulation of this process may contribute to the development of different diseases such as cancer. This higher level of regulation requires additional proteins to facilitate the correct positioning of the spliceosomal subunits on the primary transcript. Here, we reveal a role for DIDO3, one of the three *Dido* gene products, in SFPQ binding and alternative splicing. Previous studies in our laboratory established binding to H3K4Me3 through a PHD domain located at the amino terminal region of the protein. A TFS2M domain at the central part of DIDO3 promotes association with the RNAPII jaw during transcription elongation. Now, we show interaction between the carboxy-terminus of DIDO3 and SFPQ, a known splicing factor. SFPQ is a protein that associates with the polypyrimidine tract, and in particular facilitates correct U2 snRNP positioning on the 3' splice site of exons. The generation of a *Dido* mutant lacking DIDO3 while preserving the other two isoforms suppressed binding of SFPQ to RNA and increased skipping for a large general group of exons. Exons containing SRSF1 recognition sequences however were included more efficiently. Alternative splicing was also studied in the context of another *Dido* mutant. The deletion of the amino terminal region of DIDO3, too, resulted in alternatively spliced exons, although the observed splicing defects were milder. In this mutant, downstream T-rich regions, associated with RNA polymerase II pause sites, facilitated the inclusion of their upstream exons.

Together, our results indicate that the DIDO3-SFPQ association regulates alternative splicing. Due to its modular structure, DIDO3 could act as a bridge between RNA polymerase II and SFPQ, and thereby control the recruitment of the latter to the nascent RNA. Lack of the DIDO3-specific domain reduces SFPQ availability for the RNA and promotes the skipping of exons that are highly SFPQ-dependent. Although alternative pathways also regulate exon inclusion or skipping, we propose that the DIDO3-SFPQ axis in particular has evolved to expand alternative splicing regulation and maintain RNA splicing efficiency in mammals.

# RESUMEN

# RESUMEN

El *splicing* alternativo es un mecanismo fundamental que permite la generación de múltiples isoformas de proteínas a partir del mismo ARN mensajero. Aunque este es un proceso necesario en todos los organismos eucariotas desde levaduras hasta mamíferos, el *splicing* alternativo es más común en eucariotas superiores como los vertebrados. Casi el 95% de los genes de mamíferos hacen *splicing* alternativo, por lo que fallos en la regulación de este proceso promueven la aparición de muchas enfermedades como cáncer. Este alto nivel de regulación requiere por tanto el uso de proteínas adicionales para facilitar el correcto posicionamiento de las subunidades del *spliceosoma* al transcrito primario. En este proyecto desvelamos el papel de DIDO3, uno de los tres productos del gen *Dido*, en la unión con SFPQ y el *splicing* alternativo. Estudios previos establecieron que DIDO3 se une a la H3K4Me3 a través de un dominio PHD localizado en el extremo amino terminal de la proteína, y a la RNAPII durante la elongación de la transcripción a través de un dominio central TFS2M. En esta tesis mostramos la interacción entre el extremo carboxi-terminal de DIDO3 y un factor de *splicing* conocido como SFPQ. Esta proteína se asocia al "polypyrimidine tract" y en concreto, facilita el correcto posicionamiento del U2 snRNP en los sitios 3' de aceptación de algunos exones. La generación de un mutante de *Dido* carente de la isoforma DIDO3, pero manteniendo las otras dos isoformas, suprime la unión de SFPQ con el ARN y aumenta el *skipping* de un gran grupo de exones. Sin embargo, los exones que contienen secuencias de reconocimiento de SRSF1 se incluyen de una forma más eficiente. Estudiamos además el *splicing* alternativo en el contexto de otro mutante de *Dido*, uno carente del extremo amino-terminal de la proteína. Aunque con esta deleción también se observaron defectos en el *splicing* alternativo, éstos eran más moderados que los del mutante del carboxi-término. En este mutante, regiones ricas en Timina localizadas *downstream* de los exones y asociadas a pausas de la ARN polimerasa II, facilitaban la inclusión de sus exones *upstream*.

Todos estos resultados indican que la asociación DIDO3-SFPQ regula el *splicing* alternativo. Gracias a su estructura modular, DIDO3 podría actuar como un puente entre la ARN polimerasa II y SFPQ, y de este modo controlar el reclutamiento de este último al ARN naciente. La carencia del dominio específico de DIDO3 reduce la disponibilidad de SFPQ para unirse al ARN y promueve el *skipping* de los exones dependientes de SFPQ. Aunque hay otras vías que regulan el *splicing* alternativo, proponemos que el eje DIDO3-SFPQ en particular ha evolucionado para aumentar el nivel de regulación del proceso y mantener la eficiencia en el *splicing* del ARN en mamíferos.

# ABBREVIATIONS

# ABBREVIATIONS

| | |
|---|---|
| 4SU | 4-thiouridine |
| 5-mers | Pentamers |
| AS | Alternative Splicing |
| ATP | Adenosine triphosphate |
| BLAST | Basic Local Alignment Search Tool |
| Bp | Base pairs |
| BSA | Bovine Serum Albumin |
| CBC | Capping Binding Complex |
| CC | Coiled-coil |
| cDNA | complementary DNA |
| CPA | Cleavage and Polyadenylation |
| CPF | Cleavage and Polyadenylation Factors |
| CTD | Carboxy Terminal Domain |
| CTL | Control |
| DMEM | Dulbecco's Modified Eagle Medium |
| DMSO | Dimethyl sulfoxide |
| DSB | Double strand breaks |
| DTT | Dithiothreitol |
| ESC | Embryonic Stem Cells |
| ESE | Exonic splicing enhancers |
| ESRs | Exonic Splicing Regulatory sequences |
| ESS | Exonic splicing silencers |
| FA | Formaldehyde |
| FBS | Fetal Bovine Serum |
| FRT | Flp Recombination Target |
| GFP | Green Fluorescent Protein |
| GST | Glutathione-S-transferase |
| GTP | Guanosine triphosphate |
| H3K4Me3 | Histone 3 Lysine 4 Trimethylated |
| HEK | Human Embryonic Kidney |
| HeLa-FLP | HeLa Flp-In™ T-Rex™ |
| hnRNP | Heterogeneous nuclear Ribonucleoprotein |

| | |
|---|---|
| iCLIP | Individual - nucleotide resolution Crosslinking and Immunoprecipitation |
| IGK | Isoginkgetin |
| IP | Immunoprecipitation |
| iPSC | Induced pluripotent stem cells |
| ISE | Intronic splicing enhancers |
| ISS | Intronic splicing silencers |
| MEF | Mouse Embryonic Fibroblasts |
| min | Minutes |
| mRNA | Messenger RNA |
| NCL | Nucleolin |
| NPM | Nucleophosmin |
| nt | Nucleotides |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PAR-CLIP | Photoactivatable Ribonucleoside - Enhanced Crosslinking and Immunoprecipitation |
| PAS | Polyadenylation Site |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase Chain Reaction |
| PHD | Plant Homeodomain |
| Poly(A) | Polyadenylation |
| PPT | Polypyrimidine tract |
| PTM | Post-translational modifications |
| qPCR | Quantitative PCR |
| RBD | RNA Binding Domain |
| RBP | RNA Binding Protein |
| RNA seq | RNA sequencing |
| RNAPII | RNA Polymerase II |
| RPE | Retinal pigment epithelium |
| RQ | Relative Quantification |
| RRM | RNA Recognition Motif |
| RT-PCR | Reverse Transcriptase PCR |

# ABBREVIATIONS

| | |
|---|---|
| SD | Standard Deviation |
| SDS | Sodium Dodecyl Sulfate |
| SF | Splicing Factor |
| sgRNA | RNA guides |
| snRNP | small nuclear Ribonucleoprotein |
| SRSF | Serine Arginine Splicing Factor |
| SS | Splice Sites |
| TFS2M | Transcription elongation Factor S-II subunit M |
| TSS | Transcription Start Site |
| U1 | U1 snRNP |
| U2 | U2 snRNP |
| WB | Western Blot |

# INTRODUCTION

## 1. The splicing process

RNA metabolism is a highly regulated process acting as a linker between the coding information (DNA) and the final executors (proteins). This process encompasses the synthesis, processing, export and degradation of the RNA molecules. While the synthesis is based on the transcription of DNA into RNA, the processing comprises the post-transcriptional modifications of this newly synthesized RNA molecule. The mechanism includes capping, splicing and polyadenylation steps, to get in the end a mature messenger RNA. Although, in the past, transcription and processing were believed to be independent, we know now that this concept is not entirely correct. Both processes are frequently coupled and occur simultaneously (Beyer and Osheim 1988, Neugebauer 2002).

One of the mechanisms implicated in RNA maturation is known as RNA splicing, and is a necessary step for the expression of most genes in eukaryotes (Iannone and Valcarcel 2013, Kornblihtt, Schor et al. 2013). Pre-messenger RNA splicing is based on the removal of intronic sequences from the primary transcripts. The remaining exons are joined together to produce mature messenger RNAs (mRNAs) ready to be translated into proteins. Alternative splicing is a very flexible process that allows higher vertebrates to evolve by promoting proteome expansion and gene expression regulation (De Conti, Baralle et al. 2013).

The splicing process is highly conserved among organisms. If we visualize an RNA molecule as a group of randomly arranged nucleotides, then evolution had to designate specific sequences to show the splicing machinery where a section begins and ends. When this section is an intron, we refer to this mechanism as "intron definition model". On the other hand, when an exon needs to be recognized, we talk about "exon definition model". These two models of splicing correlate with the length of each section. While higher eukaryotes such as vertebrates are characterized by long introns and short exons, lower eukaryotes, for example yeasts, contain large exons separated by small introns. These differences make yeast more likely to use the intron definition model, whereas vertebrates would probably adapt its recognition mechanism to short exons and hence make use of an exon definition model (Amit, Donyo et al. 2012, De Conti, Baralle et al. 2013).

### 1.1. Spliceosome and RNA recognition sequences

The machinery responsible for exon/intron recognition is the spliceosome, a multi-megadalton complex formed by five uridine-rich small nuclear ribonucleoproteins (U1, U2, U4, U5 and U6 snRNPs). Each snRNP consists of multiple proteins and an RNA scaffold. These proteins assemble in a 1:1 stoichiometry to form the functional spliceosome (Wahl, Will et al. 2009, Will and Luhrmann 2011). In addition to the snRNPs that form the core spliceosome, around 150 non-snRNP proteins may help the process (Cvitkovic and Jurica 2013). A typical human gene contains around 8 introns that need to be processed to produce a mature messenger RNA molecule. While only one RNA polymerase II (RNAPII) is needed to transcribe an entire gene, multiple spliceosomes are required to remove the introns. Spliceosome assembly is based on the proper recognition of specific sequences at the exon/intron boundaries, known as 5' and 3' splice sites (SS). These sites, together with the branch site and the polypyrimidine tract (PPT) located near the 3'SS, contain consensus sequences that are recognised by the snRNPs and auxiliary factors such as U2AF35 and U2AF65 (**Figure 1**) (De Conti, Baralle et al. 2013, Herzel, Ottoz et al. 2017). Despite what we might think, these regions are poorly conserved in metazoans (Wahl, Will et al. 2009).

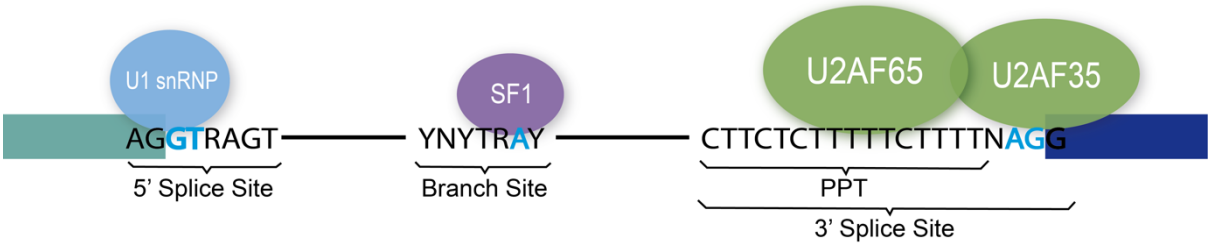



**Figure 1. Early steps of splicing factor recruitment to pre-mRNA molecules**. Small nuclear ribonucleoproteins (snRNPs) such as U1 snRNP and auxiliary factors (SF1, U2AF35 and U2AF65) bind to consensus sequences at the intron and exon boundaries. Adapted from (Kornblihtt, Schor et al. 2013).

During decades, two different models of spliceosome assembly were studied: (1) the step-wise assembly model, where individual snRNPs are assembled sequentially on a pre-mRNA, and (2) the penta-snRNP or supra-spliceosome model, which proposes that a preformed spliceosome complex containing all the snRNPs binds the pre-mRNA (Rino and Carmo-Fonseca 2009). Past studies have defended the penta-snRNP model, as they found the five snRNPs to be associated in the absence of pre-mRNA (Konarska 1988, Hausner 1990, Wassarman 1992). However, "step-wise assembly" is the most accepted model nowadays (Behzadnia, Hartmuth et al. 2006). Different publications have shown that the U1 and U2

snRNP can bind to pre-mRNA before associating with the tri-snRNP, or even in the absence of a U5 snRNP (Gornemann, Kotovic et al. 2005, Lacadie and Rosbash 2005).

According to the "step-wise assembly" model, U1 snRNP first recognises the canonical sequence of the 5'SS while the U2AF heterodimer (U2AF35 and U2AF35) binds to the PPT and 3'SS (**Figure 2**). The U1 interaction is stabilized by SF1 binding at the branch site. These proteins together with pre-mRNA form the E complex. When the U2 snRNP stably binds to the branch site, SF1 is displaced and U2AF auxiliary factors are removed from the 3'SS, converting the E complex into the A complex. This step facilitates association of the preassembled U4/U6-U5 tri-snRNP to form the B complex. This complex undergoes several conformational and compositional rearrangements that provoke dissociation of the U1 and U4 snRNP. When the B complex is catalytically activated, two transesterification steps generate the RNA lariat and fuse the exons. The first catalytic step is based on a nucleophilic attack on the branch site adenosine by the 5'SS, promoting the transition from B to C complex. Further rearrangements stimulate the second transesterification reaction in which the 3'SS is cleaved, RNA lariat is removed, and exons are joined (Rino and Carmo-Fonseca 2009, Wahl, Will et al. 2009).

### 1.2. Alternative Splicing

The possibility to generate multiple protein isoforms by the same pre-mRNA sequence is based on one fundamental mechanism known as Alternative Splicing (AS). It was discovered together with splicing itself (Chow LT, RE et al. 1977), but only AS allows metazoans to extend proteome diversity without significantly increasing the number of genes. Different deep sequencing analyses estimate that nearly 95% of mammalian genes undergo alternative splicing (Wang, Sandberg et al. 2008, Barash, Calarco et al. 2010). Thus, mis-regulation of this process may lead to aberrant protein production and could contribute to multiple diseases such as cancer (Auboeuf, Carmo-Fonseca et al. 2012, Liu and Cheng 2013, Zhang and Manley 2013, Oltean and Bates 2014, Singh and Eyras 2017), neuronal pathologies (Wan and Dreyfuss 2017), or muscular disorders (Singh and Eyras 2017, Wan and Dreyfuss 2017). Indeed, a quick search for alternative splicing and cancer in Pubmed yields more than seven thousand publications, highlighting the close relation of this process with cell development and survival.

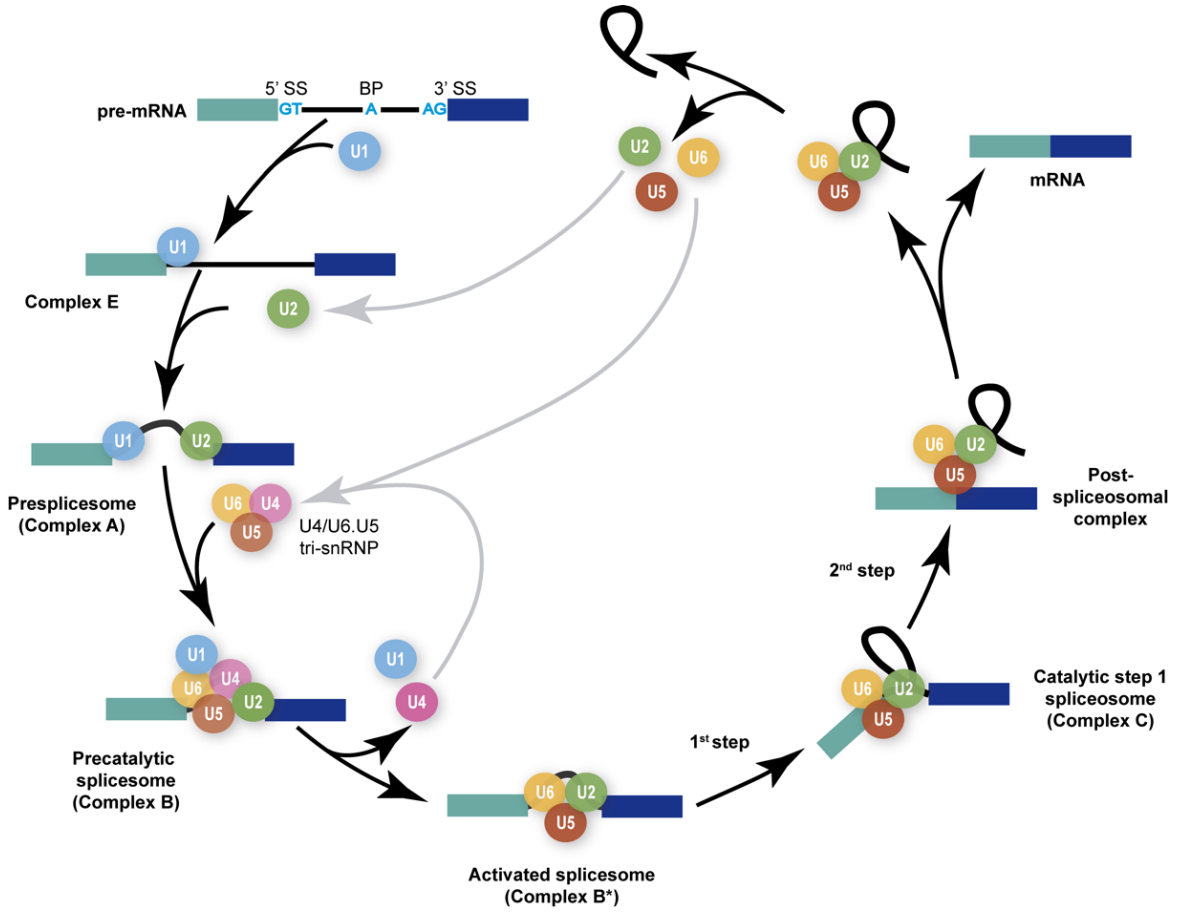



**Figure 2. Step-wise model of spliceosome assembly.** Core spliceosome proteins bind to intronic sequences in a very dynamic process. All of them are recycled when the corresponding step is completed. Adapted from (Wahl, Will et al. 2009).

Alternative splicing is more prevalent in higher than in lower eukaryotes. As compared to the invertebrates, the vertebrates show an increased amount of genes that undergo AS (Keren, Lev-Maor et al. 2010). Interestingly, although yeast is an animal model for studying the splicing process, AS events are very rare in this organism (Howe 2003). However, AS may vary considerably between different vertebrates. Considering the high conservation of DNA sequences in vertebrates, splicing in mouse and human only share a low degree of similarity (Barbosa-Morais., Irimia. et al. 2012). There are different types of AS events (**Figure 3**). Whereas intron retention is the most common event found in lower metazoans (plants and even in fungi and protozoans), exon skipping occurrence increases along the phylogenetic tree together with organism complexity (Kim, Goren et al. 2008). As mentioned in the previous section, exon and intron length is a major determinant in the splicing mechanism. These differences are also important for defining AS event types. For example, the length of the upstream intron has a higher influence on exon selection than the

downstream intron (Fox-Walsh, Dou et al. 2005). In addition, in the human genome, exons flanked by long introns are more likely to undergo exon skipping (Fox-Walsh, Dou et al. 2005). Although all of these factors can influence AS patterns, a huge variety of proteins known as splicing factors are the real regulators of this process.

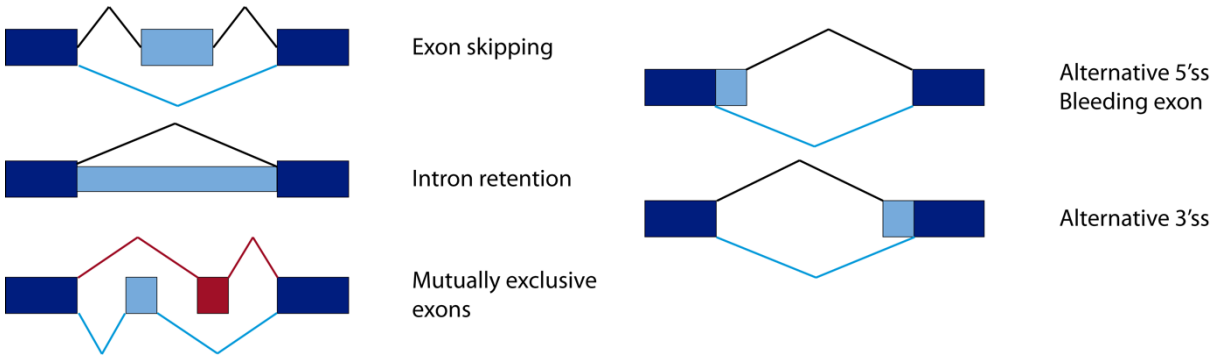



**Figure 3. Types of alternative splicing events.** Dark blue boxes represent constitutive exons. Cyan and red boxes depict alternative exons or introns. Black lines show constitutive splicing events while cyan lines illustrate alternative splicing events.

### 1.2.1. non-snRNP splicing factors

The non-snRNP splicing factors are proteins associated with the splicing process that regulate the recognition of splice sites by the spliceosome machinery. These proteins, too, favour the recognition of alternative splice sites and promote the removal of canonical exons or the inclusion of unusual ones. Thus, non-snRNP splicing factors are important to define alternative splicing events such as exon skipping o inclusion, even though they do not participate in the splicing reactions.

Nowadays, one of the most widely accepted hypothesis about how splicing factors regulate alternative splicing is based on *cis*-regulatory sequences and *trans*-acting factors. It has been described that splicing is controlled by these regulatory sequences in the pre-mRNA, comprised by exonic splicing enhancers (ESE), exonic splicing silencers (ESS), intronic splicing enhancers (ISE) and intronic splicing silencers (ISS) (House and Lynch 2008, Kornblihtt, Schor et al. 2013). These sequences are recognised by *trans*-acting factors such as the Serine-Arginine Splicing Factor (SRSF) family, heterogeneous nuclear ribonucleoproteins (hnRNPs) (Busch and Hertel 2012), as well as tissue-specific factors like PTB (Romanelli, Diani et al. 2013), NOVA (Jelen, Ule et al. 2007) and FOX (Lee, Tang et al. 2009). These factors bind to spliceosomal components and may have either activating or

inhibitory effects on the use and recognition of the 3' and 5' SS (Kornblihtt, Schor et al. 2013). U1 snRNP recognition of the 5'SS does not require additional helper proteins because it interacts directly with the RNA (Spiluttini, Gu et al. 2010). However, U2 snRNP placement on the 3'SS seems to be guided by accessory splicing factors, among others the U2AF heterodimer, SRSF family or SFPQ proteins (Ruskin, Zamore et al. 1988, Patton, Porro et al. 1993, House and Lynch 2008).

Nonetheless, hundreds of splicing factors are implicated in AS regulation in different ways. SFPQ, also known as PSF (polypyrimidine tract-binding protein-associated splicing factor), belongs to a fascinating group of factors which regulate both splicing and transcription and additionally interact with components of their machineries. Aside from SFPQ, $p54^{nrb}$/NONO, Prp40, FBP11 and TCERG1 have also been included in this category (Montes, Becerra et al. 2012). Both SFPQ and $p54^{nrb}$/NONO are ubiquitously expressed in mammalian tissues. These proteins are involved in gene and transcriptional regulation, RNA processing, DNA repair and even snRNA nuclear export (Knott, Bond et al. 2016). Both interact with the spliceosome (Peng, Dye et al. 2002, Peng, Hawkins et al. 2014) and also with specific pre-mRNA sequences (Buxade, Morrice et al. 2008). Besides, the interaction between SFPQ and the hypo and hyperphosphorylated RNAPII carboxyterminal domain (CTD) has been published (Emili, Shales et al. 2002). However, more recent studies have found no connexion of SFPQ with the polymerase (Das, Yu et al. 2007).

The *SpliceAid-F* database, which allows searching for published splicing factors, was released in 2013. This website contains a long list that covers most splicing factors found in humans, including their RNA binding sites, their different mutations, and related diseases (Giulietti, Piva et al. 2013). To directly recognize RNA molecules, most of these proteins contain well-known RNA binding domains (RBDs), such as RNA recognition motifs (RRM), KH-homology domain (KH), double-stranded RBD (dsRBD), zinc fingers (Znf) and DEAD box helicase domains (Lunde, Moore et al. 2007). Furthermore, recent studies have also found novel RBDs in flexible and disordered regions of RNA binding proteins (RBPs). These domains include arginine (R) and serine (S) (RS) repeats, arginine and glycine (G) (RG[G]-rich) repeats and lysine (K) or arginine rich sequences (K/R basic patches), which may be involved in both specific and non-specific interactions with messenger RNA. The same domains affect

RNA granule formation and participate in co- and post-transcriptional RNA metabolism (Castello, Fischer et al. 2016, Jarvelin, Noerenberg et al. 2016).

### 1.2.2. RNA Polymerase II CTD

Although splicing factors are major regulators of alternative splicing, these proteins not only are recruited by the nascent RNA itself but also via other proteins. The most extensively studied example of this protein-protein recruitment is the RNA polymerase II appendage known as CTD. This carboxy-terminal domain (CTD) is situated on the RNAPII large subunit and comprises a repetition of the heptapeptide sequence (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7). The number of repeats is highly conserved in every species (52 repeats in humans and 26 repeats in yeast) (Corden 1990). Already many years ago, the RNAPII CTD was implicated in coupling mRNA processing to transcription (McCracken, Fong et al. 1997) and its phosphorylation status was shown to have a strong effect on mRNA splicing (Hirose, Tacke et al. 1999). Nowadays, the way in which the post-translational modification of the CTD (Ser2-P, Ser5-P, Ser7-P or Thr4-P and Tyr1-P) could affect transcription and RNA processing has been widely studied (Nojima, Gomes et al. 2016). In budding yeast, for example, Ser5 and Ser7 phosphorylation occur early in transcription, during initiation and elongation over the first intron. On the other hand, Ser2, Thr4 and Tyr1 are phosphorylated during transcription of the remaining exons and termination (Herzel, Ottoz et al. 2017).

The impact of splicing inhibition on Ser2 phosphorylation has been studied in a recent publication (Koga, Hayashi et al. 2015). Here, a bidirectional regulation between splicing and RNAPII CTD phosphorylation patterns was proposed. Nevertheless, the contribution of the RNAPII CTD and splicing factors to AS outcome is still controversial. Several models with regard to AS regulation have been proposed over the years, among which the kinetic coupling model and the recruitment coupling model are the most widespread.

### 1.2.3. Models associated with alternative splicing

That splicing is mainly co-transcriptional is a widely accepted statement nowadays. Nevertheless, clear evidence for post-transcriptional splicing has also been published

(Tardiff, Lacadie et al. 2006). Various experiments identified unspliced but polyadenylated transcripts in the cell nucleus, suggesting that splicing can be done at different time points, maybe to regulate the expression of certain genes (Bhatt, Pandya-Jones et al. 2012, Boutz, Bhutkar et al. 2015).

Still, many models regarding the AS process have been elaborated and most of them assume that splicing is carried out co-transcriptionally. Although this concept first appeared at the end of twentieth century, many current publications still try to decipher the precise moment when splicing begins. Karla Neugebauer's group recently published a study in yeast, in which they measured that moment. The splicing onset occurs when RNAPII is only ±26 nucleotides (nt) downstream the 3'SS. At this moment, 15 nt of nascent RNA are protected by RNAPII and 9 nt are bound by the spliceosome (Oesterreich, Herzel et al. 2016). Thus, intron removal starts as soon as the 3'SS emerges from RNAPII and its recognition by spliceosome proteins is possible. Following these findings and the co-transcriptional splicing concept, two main regulation models have been proposed, the recruitment and the kinetic coupling model.

#### 1.2.3.1. The recruitment coupling model

The ***recruitment or spatial coupling model*** is based on the recruitment of processing factors to the site of transcription. This function is mostly carried out by the RNAPII CTD (**Figure 4**). The CTD has a key role in functionally linking these processes by acting as a "landing pad" for RNA processing factors. Not only splicing factors but also capping and 3' end processing factors bind to the CTD during transcription (Custodio and Carmo-Fonseca 2016). Indeed, the capping enzyme and the cleavage/polyadenylation (CPA) complex directly interact with Ser5-P (Schroeder, Schwer et al. 2000, Schwer and Shuman 2011) and Ser2-P (Licatalosi, Geiger et al. 2002, Ahn, Kim et al. 2004) CTD isoforms, respectively.

Although several studies claim that splicing factors associate with the RNAPII CTD, whether these interactions are direct or RNA-mediated remains to be confirmed (Kornblihtt, Schor et al. 2013, Saldi, Cortazar et al. 2016). Besides, it was published that CTD is not sufficient to promote co-transcriptional splicing or capping *in vivo*, indicating that other RNAPII specific subunits or associated factors are necessary (Natalizio, Robson-Dixon et al. 2009). One example is the mediator complex, which usually binds to enhancers and

promoters on the DNA. This protein complex can recruit splicing factors such as hnRNP L, SF3B or Eval1 (Huang, Li et al. 2012). Based on this kind of observations, a new field of study that considers the chromatin template itself as another source of splicing factor recruitment is getting stronger in the last few years. For example, histone chaperones and chromatin remodelers such as SWI/SNF can interfere in AS, by being incorporated into hnRNP particles on nascent transcripts (Tyagi, Ryme et al. 2009). Finally, even covalently modified histones may somehow control splicing changes. It has been published that H3K36 methylation can recruit polypyrimidine tract-binding protein (PTBP) and SRSF1 to the chromatin template with the help of other histone binders such as MRG15 and Psip1 (Luco, Pan et al. 2010, Pradeepa, Sutherland et al. 2012). Another example of an adaptor molecule is CHD1, which not only interacts with histone 3 trimethylated on lysine 4 (H3K4Me3) through its chromodomain but also binds the U2 snRNP subunit SF3A. When CHD1 or H3K4Me3 levels decrease, the association between U2 snRNP and the pre-mRNA is impaired and AS patterns are affected (Sims, Millhouse et al. 2007).

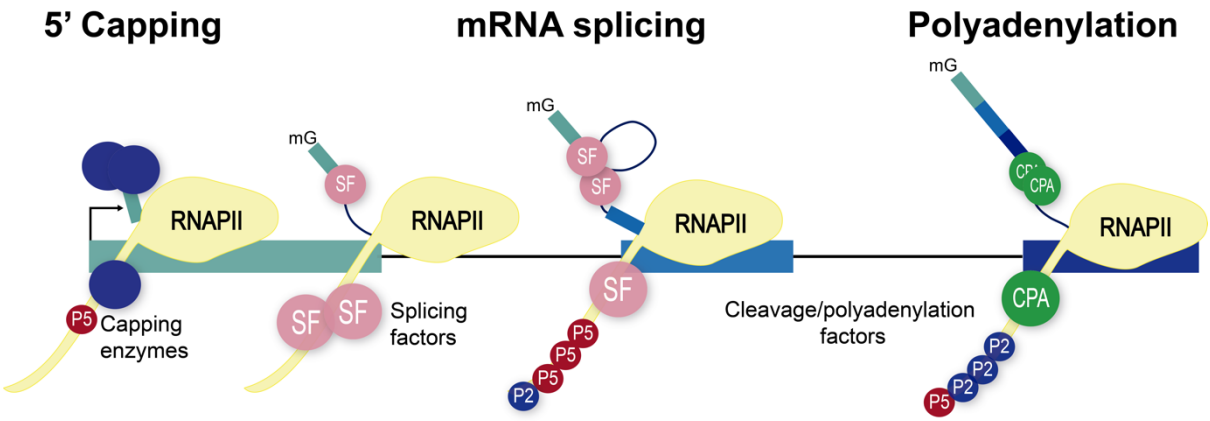



**Figure 4. Schematic representation of recruitment coupling model.** Not only splicing factors but also capping enzymes or CPA factors are recruited through the RNAPII CTD during the transcription process.

### 1.2.3.2. The kinetic coupling model

The second most studied model is **kinetic coupling**, which is based mainly on how RNAPII elongation rate and splice site strength influence AS outcome. When RNAPII transcribes a gene, introns and exons and therefore their splice sites and regulatory sequences, are exposed at different time points. On the other hand, splice sites can be weak or strong, based on how far their sequences diverge from the consensus sequence. Thus,

small variations in these two major factors could make the difference between an exon inclusion and an exon skipping event.

A strong splice site is easier to recognize by the spliceosome than a weak splice site. In a typical transcript, weak sites are located further upstream than strong sites. Thereby, when the elongation rate of RNAPII is fast, both sites are exposed at nearly the same time. The equal exposition will favour the recruitment of splicing proteins to the strong site instead of the weak site, resulting in exon skipping (**Figure 5, left**). However, sequences that induce RNAPII pausing (Roberts, Gooding et al. 1998, Carrillo Oesterreich, Preibisch et al. 2010, Henriques, Gilchrist et al. 2013) or drugs such as Camptothecin that reduces elongation rate (Boireau, Maiuri et al. 2007, Ip, Schmidt et al. 2011) can facilitate the inclusion of exons with weak splice sites (**Figure 5, right**). At the same time, transcription factors that promote elongation (Nogues, Kadener et al. 2002) or drugs that induce an open chromatin state contribute to exon skipping (Batsche, Yaniv et al. 2006, Schor, Rascovan et al. 2009). This mechanism is also known as the "**window of opportunity**" model or the "**first come, first served**" model (Aebi and Weissmann 1987), because it describes a competitive system.

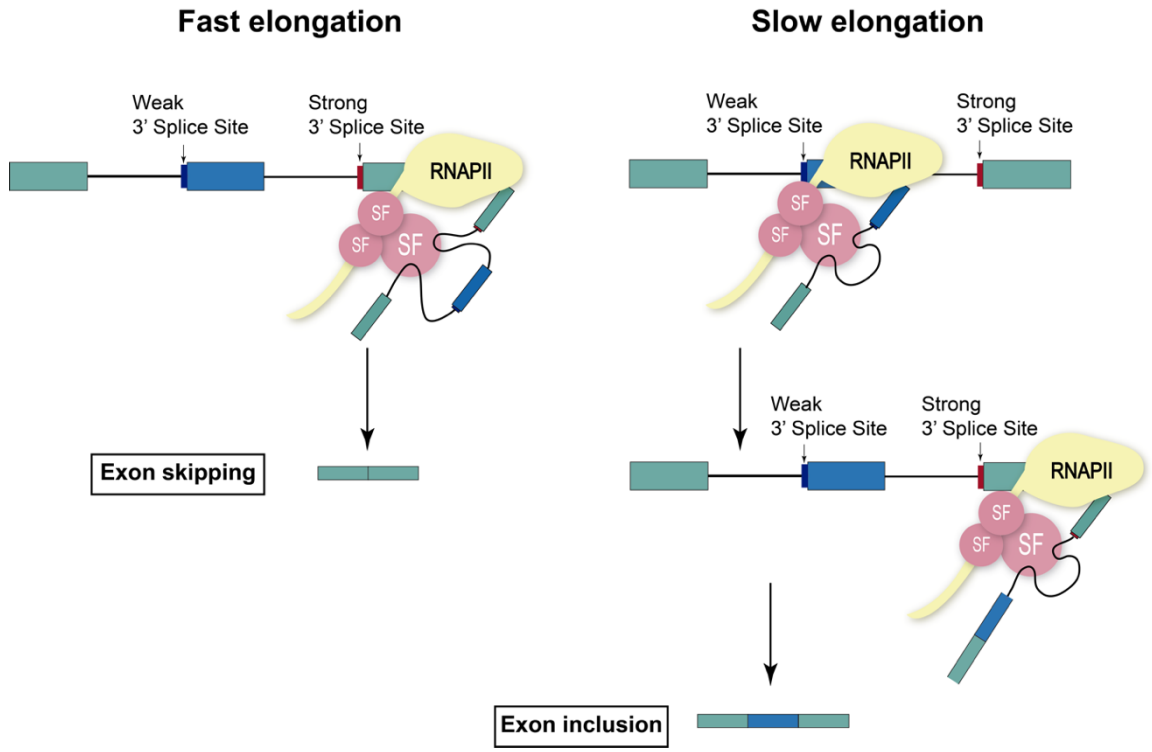



**Figure 5. Effects of the kinetic coupling model on alternative splicing.** When two 3' splice sites have different strength, RNAPII elongation rate is a major influencer of exon skipping or inclusion. Adapted from (Kornblihtt, Schor et al. 2013).

Nevertheless, the kinetic coupling model is not as simple as it seems. In some contexts, a slow RNAPII rate may favour the binding of negative regulatory splicing factors, which impair spliceosomal protein recruitment and promote the skipping of canonical exons (Dutertre, Sanchez et al. 2010, Dujardin, Lafaille et al. 2014). Another factor implicated in co-transcriptional splicing is nucleosome occupancy. Because of the differential nucleotide content found in exons and introns, nucleosomes are more densely packed on exonic than intronic regions (Schwartz, Meshorer et al. 2009, Spies, Nielsen et al. 2009). While exons are essentially rich in G-C nucleotides, A-T content is higher in intronic regions (Amit, Donyo et al. 2012). In addition, nucleosomes are positioned more stably in exons flanked by weak splice sites than in exons with strong splice sites (Tilgner, Nikolaou et al. 2009). These data suggest that nucleosomes, and thereby chromatin structure, may have an effect on alternative splicing (Schwartz, Meshorer et al. 2009, Tilgner, Nikolaou et al. 2009). Different studies have related nucleosome occupancy with RNAPII elongation rate as well (Bondarenko, Steele et al. 2006, Bintu, Ishibashi et al. 2012). Thus, when the polymerase II encounters a nucleosome, this generates a transient pause that marks the beginning of each exon (Schwartz, Meshorer et al. 2009). This delay will provide more time to the splicing machinery for the recognition of weak splice sites.

In addition to the epigenetic factors described above, different sequences in DNA and RNA can promote an RNAPII pausing state. Several studies have associated AT-rich sequences with RNA polymerase pausing in bacteria, yeast, and mammals (Artsimovitch and Landick 2000, Palangat, Hittinger et al. 2004, Henriques, Gilchrist et al. 2013). Others have proposed G-rich termination elements as poly(A) proximal pause sequences that promote efficient transcriptional termination (Ashfield, Enriquez-Harris et al. 1991, Gromak, West et al. 2006). A well-studied example of pausing elements is the so-called backtracking. This phenomenon is based on the reversible movement of RNAPII along DNA and RNA when the nascent transcript is displaced from the polymerase active site. This effect prolongs specific paused states without depolymerizing the transcript (Sahoo and Klumpp 2013). The RNAPII backtracking can be rescued by the TFIIS transcription factor. This protein interacts with the active site of the enzyme and promotes the activity of an intrinsic endonuclease to cut and excise the backtracked RNA. Once the RNA is cleaved, the 3' end realigns at the active site and transcription can continue (Cheung and Cramer 2011, Nudler 2012). TFIIS protein thus

increases RNAPII elongation rate by helping to overcome backtracking (Nudler 2012, Pinskaya, Ghavi-Helm et al. 2014). In the end, this transcription factor influences the outcome of co-transcriptional splicing by decreasing the window of opportunity during which exons can be included.

### 1.2.4. Cooperation and competition between splicing and polyadenylation machineries

We have discussed how splicing factors, RNAPII CTD and other related proteins regulate alternative splicing and in some cases polyadenylation. However, we have not considered that spliceosome and poly(A) machineries are also capable of regulating each other. It was described long ago that the U1 snRNP can inhibit the pre-mRNA polyadenylation by interacting directly with the poly(A) polymerase (Gunderson, Polycarpou-Schwarz et al. 1998). In addition, more recent studies show that this canonical spliceosomal factor binds to dormant intronic CstF64 sites to avoid premature polyadenylation (Yao, Biesinger et al. 2012). Indeed, U1 impact on the poly(A) machinery has been widely studied over the last 10 years, specially by the group of Gideon Dreyfuss. They have postulated several times that the U1 snRNP prevents premature cleavage and polyadenylation (PCPA), in a process they named Telescripting. Telescripting is necessary to maintain long pre-mRNA transcripts (Kaida, Berg et al. 2010, Berg, Singh et al. 2012). Thus, U1 levels, apart from controlling the splicing process, has a key role in poly(A) site (PAS) usage and therefore in regulating mRNA length. The disruption of this factor promotes an extensive usage of more proximal PAS, generating mRNAs with shorter 3' UTRs. This phenomenon is characteristic of activated immune, neuronal and cancer cells (Niibori, Hayashi et al. 2007, Flavell, Kim et al. 2008, Sandberg, Neilson et al. 2008, Mayr and Bartel 2009).

In addition to shorter mRNAs, the splicing pattern of some genes is indeed affected when U1 control is compromised. It is especially important in two different splicing events, the so-called bleeding exons and mutually exclusive exons at the 3' end of genes (**Figure 6**). In both cases, the RNA processing machinery needs to decide between continuing with splicing or instead choosing the termination. In those examples, the presence of U1 suppresses PAS choice, promoting the binding of the U2 snRNP and defining the internal

exon (Berg, Singh et al. 2012). Finally, also the RNAPII transcription elongation complex, through splicing and poly(A) factor recruitment, has been associated with the regulation of both machineries (McCracken, Fong et al. 1997, Hirose, Tacke et al. 1999, Das, Yu et al. 2007).

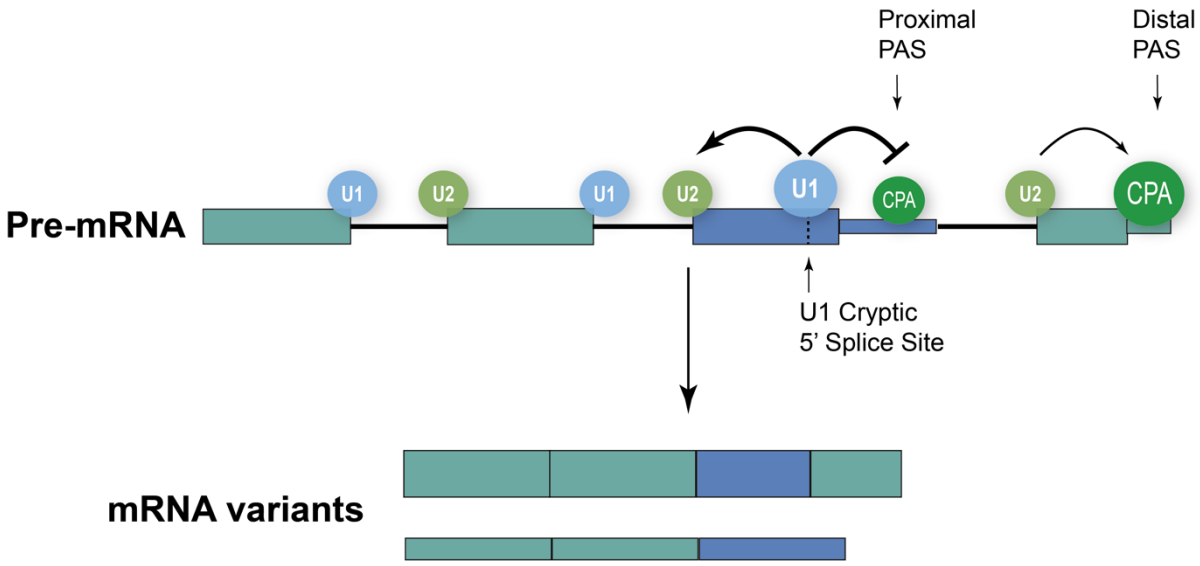



**Figure 6. Splicing (U1 and U2) and polyadenylation (CPA) machineries regulate each other**. In bleeding exons and mutually exclusive terminal exons, both machineries generate different mRNA variants. U1 negatively regulates CPA whereas it promotes U2 binding. PAS means polyadenylation site

## 2. The Death Inducer Obliterator (DIDO) protein

The death inducer obliterator (*Dido*) gene was firstly described in 1999 as *DIO-1* and was associated with the process of apoptosis in pre-B cells from murine bone marrow. Initially, only one of three isoforms was identified, and its function was related with the induction of apoptosis *in vitro* and cell death during development (Garcia-Domingo, Leonardo et al. 1999, Garcia-Domingo, Ramirez et al. 2003). However, several years later this role in apoptosis was found to be a consequence, instead of a cause, of alterations in DNA regulation produced by chromatin instability (Rojas, Sanchez-Pulido et al. 2005). Nowadays, different phenotypes have been associated with the *Dido* gene, although finding a specific central mechanism is imperative.

The *Dido* gene is located on the long arm of chromosome 20 in humans and chromosome 2 in mice. It is ubiquitously expressed and contains 16 exons. The gene codes for three different proteins generated by alternative splicing, DIDO1, DIDO2 and DIDO3. All

isoforms have a common N-terminal region and specific C-terminal parts (Figure 7) (Futterer, Campanero et al. 2005).

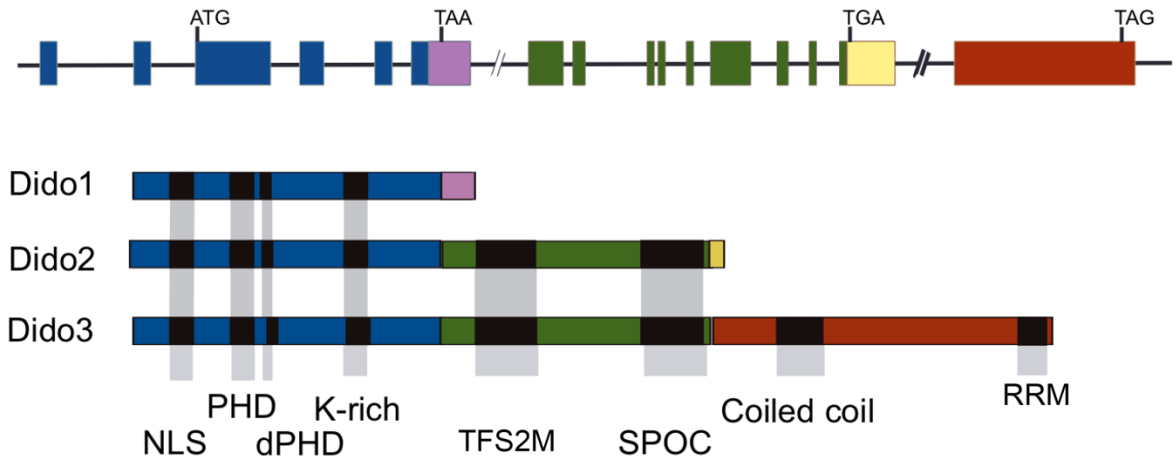



**Figure 7. Representation of the *Dido* gene and its three different isoforms with the known and predicted domains.**

As shared domains, DIDO contains a Nuclear Localization Signal (NLS), and a plant homeodomain (PHD) finger and dPHD region (short motif downstream of the PHD domain). The latter two domains have been associated with the recognition of post-translationally modified $H3K4Me3$ (Prieto, Kouznetsova et al. 2009).

The central part of the protein, only shared by DIDO2 and DIDO3, contains a transcription elongation factor S-II subunit middle (TFS2M) domain that interacts with RNAPII (Futterer, de Celis et al. 2017). The yeast homolog BYE1 uses this domain to recognize the RNAPII jaw and funnel domain (Kinkelin, Wozniak et al. 2013). The DIDO TFS2M domain corresponds to the second domain of the elongation factor TFIIS that stimulates RNAPII transcription through DNA lesions. As we mentioned in previous sections, when RNAPII pauses and backtracks during transcription, the TFIIS elongation factor binds to the enzyme and enhances RNA cleavage activity. This RNA cleavage promotes the RNAPII escape from backtracking (Cheung and Cramer 2011). In addition to the TFS2M region, DIDO2 and DIDO3 isoforms contain a SPOC (Spen Paralog and Ortholog) domain implicated in protein-protein interactions (Sanchez-Pulido, Rojas et al. 2004, Lee and Skalnik 2012). Finally, although little is known about the DIDO3-specific exon, the exon 16, it encodes for a large unstructured region characteristic for proteins of recent evolution (Haynes and Iakoucheva 2006). This kind of disordered sequences have recently been associated with RBP and are thought to mediate RNA-protein interactions (Jarvelin, Noerenberg et al. 2016). Although we have not

yet identified a specific domain, prediction programs reveal that the exon 16 contains a central coiled-coil domain and a carboxy-terminal RNA recognition motif (RRM).

Finally, with respect to the location of the three isoforms, DIDO1 distributes over the nucleus and cytosol while DIDO2 and DIDO3 are predominantly found in the nuclear compartment (Futterer, Campanero et al. 2005, Trachana, van Wely et al. 2007). Although *Dido* is expressed ubiquitously, DIDO3 expression decreases in somatic cells as compared to Embryonic Stem Cells (ESCs), whereas it is recovered in induced pluripotent stem cells (iPSC) (Gatchalian, Futterer et al. 2013).

### 2.1. DIDO as a multifunctional protein

As mentioned before, *Dido* was initially associated with apoptosis in immune cells. Since 1999, however, a wide variety of studies and mutants have helped to elucidate the real functions of the three DIDO isoforms.

The deletion of the whole *Dido* in cell lines appears to be incompatible with life. Thus, different mutants have been generated by eliminating specific DIDO domains (**Figure 8**). The truncation of the amino-terminal part of DIDO (*DidoΔNT*) affects the PHD domain. This mutation produces aneuploidy, centrosome amplification and cytokinesis defects, all of which are related with mitotic spindle alterations (Trachana, van Wely et al. 2007) . For the same reason, the lack of histone binding generates an increased number of lagging chromosomes during anaphase, and causes double strand breaks (DSBs) adjacent to centromeres (Guerrero, Gamero et al. 2010). All of these defects provoke cell division problems and chromosome instability (Martinez and van Wely 2011). Furthermore, mice carrying this mutation suffer myeloid neoplasms (Futterer, Campanero et al. 2005), and human myeloid leukaemia patients show alterations in DIDO isoform expression. Here, DIDO2 expression is increased in advanced phases of the disease (Berzoti-Coelho, Ferreira et al. 2016). Finally, using this mutant, DIDO3 has been related with cilium size by regulating HDAC6 availability at the centrosome (Sanchez de Diego, Alonso Guerrero et al. 2014).

Another *Dido* mutation, which involved the removal of the DIDO3-specific exon 16, was generated (*Dido3ΔCT)*. This exon encodes for the carboxy-terminal part of the protein and its deletion is lethal at the mouse embryonic stage 8.5 of gestation. The loss of DIDO3

produces a delay in the onset of lineage commitment at the stage of primitive endoderm, which compromises the differentiation of ESCs. This regulation of stemness genes is controlled by the DIDO PHD finger (Gatchalian, Futterer et al. 2013). In the case of *in vivo* models, the differentiation program can be recovered by crossing *Dido3ΔCT* mice with other mutants carrying the *ΔNT* deletion. These double mutant mice (*ΔCT/ΔNT)* can survive until the end of gestation but show a reduced life expectancy, brain malformations, as well as neuromuscular and behavioural alterations (Villares, Gutierrez et al. 2015).

Finally, the relationship between *Dido* and histone modifications in mitosis or meiosis is another recurrent topic in this protein's history. The phosphorylation of Histone 3 Threonine 3 and 6 ($H_3T_3Ph$ and $H_3T_6Ph$) during mitosis correlates with the exclusion from chromatin of PHD finger proteins, such as $DIDO_3$, ING1 or $TAF_3$. In the case of $DIDO_3$, the protein is expelled from $H_3K_4Me_3$ at the beginning of mitosis, when chromatin is condensed and processes like transcription are suppressed (Gatchalian, Mora Gallardo et al. 2016). This exclusion allows $DIDO_3$ to translocate to microtubules and helps to control the mitotic spindle (Trachana, van Wely et al. 2007, Gatchalian, Futterer et al. 2013). Besides, the $H_3K_4Me_3$ mark modulates $DIDO_3$ location in meiosis, where it was found at the central region of the synaptonemal complex when trimethylation diminished (Prieto, Kouznetsova et al. 2009).

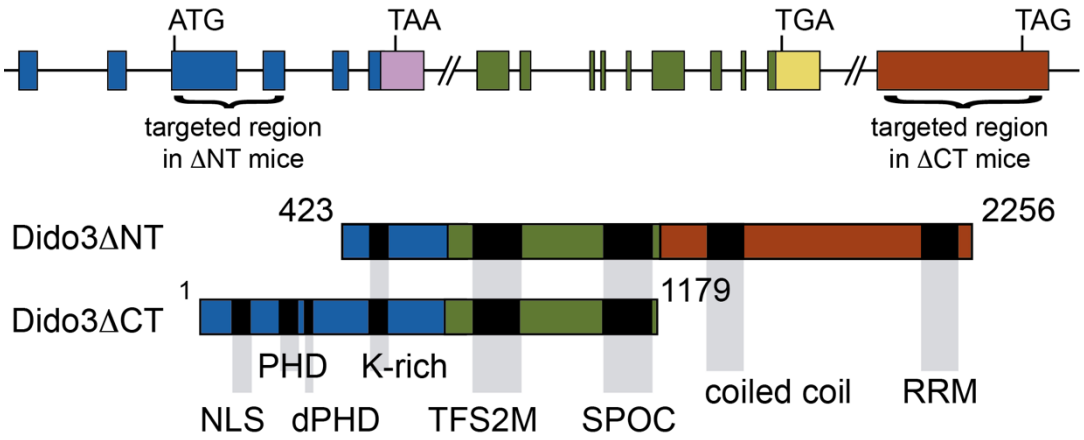



**Figure 8. *Dido* mutants lacking the N-terminal or C-terminal part of the protein.**

So, as observed, the *Dido* gene seems to be implicated in many different processes and generates a wide variety of phenotypes. These pleiotropic effects likely are related with the diversity of domains contained in the three isoforms. The distinct domains allow DIDO

to bind multiple proteins and to be part of the regulation of different pathways. However, we think that *Dido* could actually be involved in a more generic and primary process that controls and generates different outcomes. Given the interaction with RNAPII and the presence of an RRM, RNA metabolism is a good candidate of such a generic process.

### 2.2. *Dido* and evolution

The *Dido* gene and its products can be found under many different names, from DIO-1 when first described, to DATF1, C20orf158, BYE1 or PPS. Some of these terms refer to DIDO orthologs or homologs. The most studied ones are BYE1, which is expressed in yeast, and PPS found in *Drosophila melanogaster*.

The BYE1 (Bypass of Ess1) protein is a putative transcription factor identified in *Saccharomyces cerevisiae*. BYE1 contains the same domains found in DIDO1 and DIDO2 proteins. Its PHD domain interacts with $H3K4Me3$, favouring enrichment at the actively transcribed class II gene promoters. On the other hand, the TFS2M domain interacts with the RNAPII jaw (Kinkelin, Wozniak et al. 2013). The PHD domain thus may facilitate RNAPII binding by preloading active genes. Although BYE1 is not necessary for cell viability, it becomes essential for normal growth or efficient transcriptional activation under stress conditions. In these cases, a quick transcriptional response is required, and BYE1 competes with TFIIS protein to resume the transcription process (Pinskaya, Ghavi-Helm et al. 2014). The PHD, TFS2M, and SPOC domains are very well conserved from human to yeast, not only in DIDO but also in the $PHF3$ protein. The latter has been related with human glioblastoma, where its expression is dramatically reduced (Fischer, Struss et al. 2001, Struss, Romeike et al. 2001). But, although $PHF3$ contains a TFS2M domain, its mutated PHD finger impairs histone binding (Gatchalian, Futterer et al. 2013, Pinskaya, Ghavi-Helm et al. 2014).

While the amino-terminal and central domains of DIDO (DIDO1 and DIDO2) are highly conserved in eukaryotes, the C-terminal part of the protein seems to have evolved more recently. When generating a phylogenetic tree comparing DIDO, BYE1 and $PHF3$, we observed that the three proteins are classified in different groups (**Figure 9**). Whereas BYE1 is found in lower eukaryotes, DIDO and $PHF3$ are only present in vertebrates. The lampreys

seem to fall outside both groups, because they comprise a border between vertebrates and invertebrates. This way of grouping indicates that BYE1 (or DIDO2) emerged relatively early in evolution. At some point, when the vertebrates showed up and their genome duplicated (Kasahara 2007), the DIDO protein evolved individually to acquire a more specialized function. The *Dido* gene included the exon 16 (DIDO3 isoform), allowing for a new role that could not be fulfilled by the DIDO1 and 2 proteins.

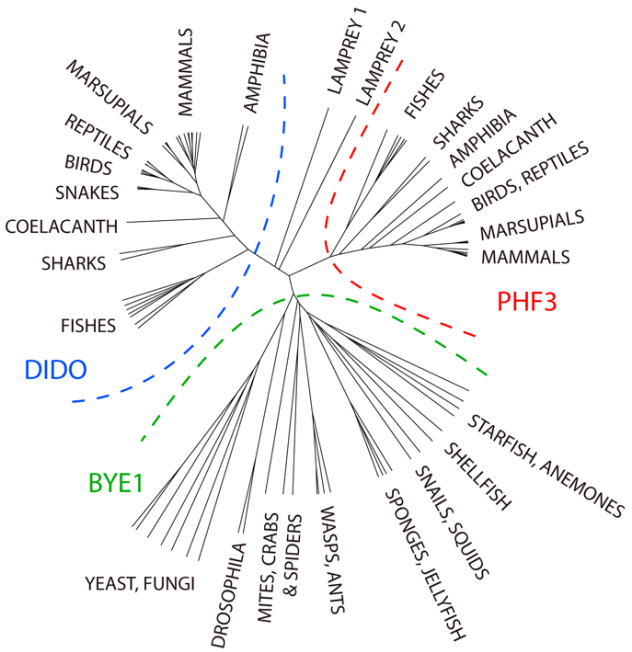



**Figure 9**. **Phylogenetic tree of DIDO homologs**. The tree was generated using ClustalΩ to align the sequences and GENEIOUS program to build the tree.

Along with genomic duplication, vertebrates evolved other important events, such as the extensive use of alternative splicing (Iniguez and Hernandez 2017). In contrast to the situation in lower eukaryotes, this process became very specialized in higher eukaryotes (Keren, Lev-Maor et al. 2010). One exception is the *D.melanogaster* sex determination gene *Sex-lethal* (*sxl*). SXL expression is controlled by an autoregulatory alternative splicing mechanism. The inclusion of the translation-terminating third exon generates a non-functional SXL protein in males, whereas this exon is always skipped producing an active coding protein in females. The system is autoregulated by SXL itself, by binding to multiple intronic sites and to different splicing factors such as the U1 snRNP protein SNF (Sans-fille). Another protein implicated in this process is the DIDO homolog, PPS, which interacts with SNF and SXL protein, as well as the unspliced *Sxl* RNA. PPS accumulates at the *Sxl* promoter, favouring the male exon skipping (Johnson, Nagengast et al. 2010).

### 2.3. DIDO and alternative splicing

Different evidences seem to relate the *Dido* gene with RNA splicing. However, only the interactions with RNAPII and H3K4Me3 have been documented (Gatchalian, Futterer et al. 2013, Kinkelin, Wozniak et al. 2013, Pinskaya, Ghavi-Helm et al. 2014, Futterer, de Celis et al. 2017). As we mentioned, the DIDO3 C-terminus is a very disordered sequence and although it contains an RRM-like domain, interacting proteins are still unknown. In order to identify possible candidates, co-purification experiments were carried out in our laboratory. Using a fusion protein formed by glutathione-S-transferase (GST) and DIDO3 without the histone-binding domain, several proteins were co-purified along with DIDO itself (**Figure 10**).

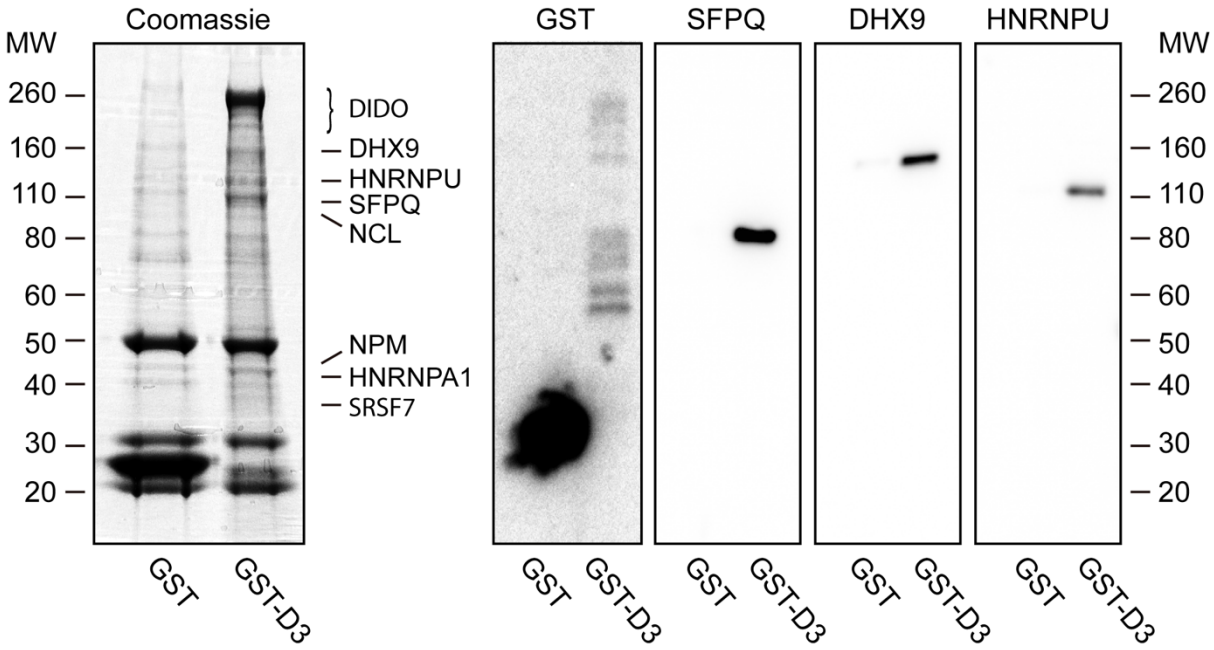



**Figure 10. GST pull down experiment using GST fused to DIDO3**. GST alone was used as a control and both constructs were purified from HEK293T cells and analysed by SDS-PAGE and Coomassie staining (on the left). Some of the interacting proteins, initially identified by mass spectrometry analysis, were confirmed by WB (on the right).

The main interacting protein was SFPQ, a well-documented splicing factor that is essential in mammals (Montes, Becerra et al. 2012,Peng, 2002 #115, Peng, Hawkins et al. 2014). In addition, proteomics analysis identified other proteins, also involved in splicing and transcription (DHX9, SRSF7, hnRNP U), and some predominantly nucleolar proteins (NCL, NPM). These evidences support the idea that DIDO3 could be implicated in RNA metabolism by regulating the alternative splicing process.

Another line of evidence derived from evolution reinforces this theory. As we pointed out in the previous section, the longest DIDO isoform is only found in organisms that make extensive use of AS but not in yeast. Remarkably, invertebrates do not produce SFPQ either. Thus, DIDO3 and SFPQ can only be found in vertebrates. Finally, the DIDO homolog PPS is implicated in the regulation of SXL alternative splicing in an organism that produces a distant SFPQ family member.

In conclusion, we already have several indications that DIDO3 is implicated in RNA splicing. We will now report a role for the DIDO3-specific domain in SFPQ binding and the mechanism followed by DIDO to regulate alternative splicing in mammalian cells.

# OBJECTIVES

# OBJECTIVES

1. **Study the DIDO3 implication in RNA metabolism through physical interactions with splicing factors**

   1.1. Identification of the DIDO isoform involved in the specific interaction with SFPQ

   1.2. Study of the subcellular location of DIDO3 and splicing factors

   1.3. Analysis of the recruitment failures associated with *Dido* mutations

   1.4. Study the interaction of DIDO3 with messenger RNA

2. **Characterization of the splicing defects associated to *Dido* mutations**

   2.1. Generation of RNA sequencing data to study the splicing alterations produced by *Dido* mutations

   2.2. Verification of the splicing defects in mouse and human models of *Dido* mutants

   2.3. Study of the DIDO3 full-length complementation in the recovery of the splicing phenotype

3. **Determine the role of the DIDO3-SFPQ axis in alternative splicing regulation**

   3.1. Analysis of the molecular mechanism underlying the DIDO3 function in alternative splicing related to SFPQ interaction

   3.2. Study of DIDO3-dependent and independent pathways in alternative splicing regulation

   3.3. Formulation of a model to explain the role of DIDO3 in alternative splicing linked to the transcription process.

# MATERIALS AND METHODS

## 1. Immortalized cell lines:

- **MEF:** Mouse embryonic fibroblast. A type of fibroblast prepared from mouse embryos. In this project MEF cell line was generated from C57BL/6 mice and immortalized using expression of human papillomavirus E6 and E7.

- **HEK293T:** Human cell line generated from HEK293 (Human Embryonic Kidney) with the viral antigen SV40 *large T.* These cells were obtained from the American Type Culture Collection (ATCC) (RRID:CVCL_0063).

- **RPE-1:** Human cell line derived from the retinal pigment epithelium and immortalized with hTERT (RRID:CVCL_4388).

- **HeLa Flp-In™ T-Rex™**: Human cell line derived from cervical cancer cells. These particular HeLa cells contain the Flp-In™ T-Rex™ system (Cat. no. K6500-01) available from Thermo Fisher Scientific.

The Flp-In™ system allows stable expression of the gene of interest by FLP recombinase-mediated integration. The T-Rex™ system provides conditional expression of the inserted gene under tetracycline or doxycycline induction. HeLa Flp-In™ T-Rex™ (HeLa-FLP) cells include a pFRT/lacZeo (from Flp Recombination Target) and a pcDNA™6/TR (from the T-REx™ System) cassettes stably integrated in their genomes. This parental cell line is resistant to Zeocin due to the pFRT/lacZeo.

MEFs and HEK293T cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS), 1% Penicillin/Streptomycin and 1% L-Glutamine. RPE-1 cells were cultured in DMEM+F12 (1:1) supplemented with 10% FBS, 1% L-Glutamine and antibiotics. Parental HeLa-FLP cells (without the inserted expression cassette) were cultured in DMEM supplemented with 10% FBS, 1% Penicillin/Streptomycin, 1% L-Glutamine and 150 µg/ml of Zeocin antibiotic to maintain the pFRT/lacZeo cassette. All cell lines were kept in cell culture incubators at 37ºC and 5% of $CO_2$.

## 2. Generation of *Dido* exon 16 mutant cells

### 2.1. Generation of MEFs ΔE16 mutants

The production of heterozygous *Dido* exon 16-floxed mice was outsourced (Ozgene Ltd, Bentley, Australia). A *loxP* site was inserted at the 3' end of *Dido* intron 15, and a neomycin cassette, flanked with FLP recombination target (FRT) sites, was inserted downstream of *Dido* exon 16. A second *loxP* site was inserted downstream, followed by a DNA sequence to reproduce the original 3' SS (See **Results**, **Figure 13**). The protein coding sequence in the targeting vector was verified. After linearization of the vector and electroporation into C57BL/6-derived Bruce4 embryonic stem cells, successful recombination was verified by southern blotting of resistant colonies. Correctly targeted clones were microinjected into recipient blastocysts for transfer into foster mothers, and chimeric offspring were crossed with C57BL/6J mice.

All subsequent work was carried out at the National Centre for Biotechnology, with approval of the local and regional Ethics Committees on Animal Experiments. Mice carrying the floxed *Dido* exon 16 were interbred with UBC-Cre-ERT2 mice (007001, Jackson Laboratory, Bar Harbor, ME) and homozygous *Dido* flox/flox *Cre* carriers were selected for further breeding. MEFs were derived and immortalized as described (Cheng, Luo et al. 2016). After treatment with 1 μM hydroxytamoxifen for 2 passages, individual clones were selected and genotyped by PCR and WB (See **Results** **Figure 13**). Homozygous deletion mutants were termed ΔE16. Primers used for genotyping are described in the **Table 1**. The ΔNT mutant generation was described before (Prieto, Kouznetsova et al. 2009).

| Primers | Sequence (5' -3') |
|---|---|
| E16 FLOXED | **Forward:** GTGTGCTGGCACATTCAGGG<br>**Reverse**: CGAGTCCGCTTCTCATCTGT |
| ΔE16 | **Forward**: TGCTCGTCAAGAAGACAGGG<br>**Reverse**: CGAGTCCGCTTCTCATCTGT |

**Table 1.** Primers for genotyping the exon 16 deletion in MEF cells.

### 2.2. Generation of ΔE16 HeLa Flp-In™ T-Rex™ cells

To generate a human cell line lacking the *Dido* exon 16, HeLa Flp-In™ T-Rex™ cells (HeLa-FLP) bearing a single FRT site were targeted with CRISPR-Cas9 (Ran, Hsu et al. 2013).

The guides used for specific deletion were designed by Amaia Talavera Gutiérrez. These guides are shown in **Table 2**. After cloning in the pSpCas9(BB)-2A-Puro vector (a gift from Feng Zhang, Addgene #62988, Watertown, MA), guides and Cas9 were delivered into HeLa-FLP cells by transient transfection. Cells were cultured with 1.5 µg/ml puromycin for 72 hours, and clones were selected after an additional 2 weeks of growth without selective pressure. Correct elimination was verified by PCR with primers shown in **Table 3** and by western blotting (See **Results Figure 27**).

| Primers | Sequence (5' -3') |
|---|---|
| hDido-E16 Guide 1 | **Forward**: CACCg**CCAAGAATTATATTCGGACG**<br>**Reverse**: AAAC**CGTCCGAATATAATTCTTGG**c |
| hDido-E16 Guide 2 | **Forward:** CACCg**AGCGAGACCGGAGGCGCGAC**<br>**Reverse**: AAAC**GTCGCGCCTCCGGTCTCGCT**c |

**Table 2.** Guides used for exon 16 deletion in HeLa-FLP cells. Letters in bold denote *Dido* sequence.

| Primers | Sequence (5' -3') |
|---|---|
| WT | **Forward:** TCCTCCTCAGGGTCGTATGG<br>**Reverse**: TAAACGTTCCCCGTGAAAGCC |
| ΔE16 | **Forward**: ATGGCTCTAGGGTCCAACGA<br>**Reverse**: TAAACGTTCCCCGTGAAAGCC |

**Table 3.** Primers for genotyping the exon 16 deletion in HeLa-FLP cells.

## 3. Plasmids and constructions

### 3.1. V5-tag plasmids for DIDO1, 2 and 3 transient expression

The three different DIDO isoforms were cloned into a vector containing the V5-tag as described (Sanchez de Diego, Alonso Guerrero et al. 2014). A pcDNA3 (Life Technologies) plasmid was used for the insertion of the mouse *Dido1*, *Dido2* and *Dido3* cDNA at the N-terminal region.

### 3.2. FRT plasmids for DIDO2 and DIDO3 stable expression

Plasmids for the stable expression of the DIDO2 and DIDO3 isoforms in the HeLa-FLP cells were generated by cloning the human cDNA in pcDNA$^{TM}$5-FRT-TO-GFP vectors. This plasmid carries the FRT sequence to stably integrate the *Dido* cDNA within the HeLa-FLP genome and an GFP reporter. To generate the DIDO3 plasmid, human *Dido3* cDNA from a

pcDNA3-HA tag vector (kindly donated by Thierry Fisher, PhD) was amplified using the following primers:

***Dido3* Forward:** 5′ TGGA GCTAGC **GACGACAAAGGCGACCCGAG** 3′

***Dido3* Reverse:** 5′ TGGA GCGGCCGC **CTAGGCCTGCGAGGCGGTGC** 3′

NheI and NotI restriction sites were inserted with the amplification to later clone the final fragment in the pcDNA5 vector. This plasmid was named as **pcDNA™5-FRT-TO-GFP-hDido3.**

To produce the DIDO2 plasmid, human *Dido2* cDNA was cloned from a pcDNA3-HA-tag plasmid (kindly donated by Thierry Fisher, PhD) into the previous pcDNA™5-FRT-TO-GFP vector. In this case, only restriction enzymes were used to insert the sequences instead of PCR amplification. The final plasmid was named as **pcDNA™5-FRT-TO-GFP-hDido2.** Both plasmids were later co-transfected with a pOG44 vector containing the Flp recombinase for an efficient stable insertion in the HeLa-FLP cells. The *Dido* inserted cassette replaces the Zeocin resistance from the parental cell lines for a Hygromycin resistance gene. Positive clones were then selected with 150 µg/ml of Hygromycin B antibiotic for 3 weeks. In contrast, negative clones that were still resistant to Zeocin did not survive in the Hygromycin supplemented culture.

### 3.3. Minigenes

To simulate the alternative splicing defect produced between *Ndufs7* and *Dazap1* genes in the ΔE16 mutant, a minigene was generated. The minigene cassette was produced in three different modules: two for *Ndufs7* and *Dazap1* mouse genes and an intermediate one including the T-rich motif downstream *Ndufs7* or a random sequence. Primers designed for the amplification of each module are shown in **Table 4**. Every primer contained a restriction site for the consecutive and proper joining between the modules. Each module was generated separately in a pBluescript II SK vector (Addgene #212205) and finally cloned in a pcDNA™5-FRT-TO_GFP plasmid.

The final vectors containing the complete minigenes were co-transfected with a pOG44 plasmid into the HeLa-FLP WT and ΔE16 cells for stable insertion. Minigenes expression was induced with 1 µg/ml of doxycycline for 72 hours.

| Primers | Sequence (5' -3') | Restriction Sites |
|---|---|---|
| ***Module 1:*** *Ndufs7* | **Forward:** GC AAGCTT **CTACTCGGTTGTTCGTGGCT**<br>**Reverse**: GA GAATTC **ACCAGAGATGTGTCTTCCGC** | HindIII<br>EcoRI |
| ***Module 2.1:*** T-rich motif | **Forward**: GA GAATTC **TATGTCAGGCCTTCGTGTCC**<br>**Reverse**: GA GGATCC **AGCTCTTCTTCACTTCGGGC** | EcoRI<br>BamHI |
| ***Module 2.2:*** Random seq | **Forward:** GA GAATTC **GACCCAGTTCTGTTCCCAGC**<br>**Reverse**: GA GGATCC **AGATCCCTTGGAGCTGGAGT** | EcoRI<br>BamHI |
| **Module 3***: Dazap1* | **Forward**: GC GGATCC **GGCCACTGCTAAGGTTCTGT**<br>**Reverse**: GA ACTAGT **GGCTTCGGGTCGATCTATGG** | BamHI<br>SpeI |

**Table 4.** Primers designed for minigene modules generation. "Random seq" means a random sequence. Restriction sites are underlined.

## 4. Immunoprecipitation of DIDO isoforms

To immunoprecipitate the three DIDO isoforms, the previously described V5-tag plasmids were transfected in HEK293T cells. For these transient transfections, $1{\cdot}10^6$ HEK293T cells were seeded in 10 cm plates. After 24 hours, 4 µg of plasmid DNA was transfected with jetPRIME liposome transfection reagent (Polyplus, Illkirch, France) according to the manufacturer´s instructions. Cells were harvested after 72 hours and subjected to immunoprecipitation with 2 µg of Rabbit polyclonal anti-V5 antibody (ab9116; RRID:AB_307024). DIDO and interacting proteins were pulled down using Protein A Dynabeads™ (Thermo Fisher Scientific Inc.) and finally eluted with 1X SDS Sample Buffer (60 mM Tris-HCl pH 6.8, 10% Glycerol, 2% SDS, 50 mM DTT, 0.05% Bromophenol). After resolving proteins on 8% SDS-PAGE, DIDO and SFPQ were visualized by western blotting using a mouse anti-V5 antibody and a rabbit anti-SFPQ, respectively. All antibodies used are shown in **Table 5**.

| Primary antibodies | Dilution | Supplier company | Catalog number |
|---|---|---|---|
| anti-V5 | 1:1000 | Abcam | ab27671, RRID: AB_471093 |
| anti-SFPQ | 1:1000 | MBL International | RN106PW |

**Table 5.** Antibodies used for membrane development in the immunoprecipitation

## 5. Immunofluorescence analysis

For immunofluorescence experiments, cells were grown until 70% confluent, rinsed in PBS, fixed in PBS containing 3.7% formaldehyde (FA), and permeabilized in PBS with 0.5%

NP40. After blocking in PBS containing 2% bovine serum albumin (BSA) and 0.05% Tween20 (1 h, room temperature), cells were incubated with primary antibodies (1 h, room temperature). Later, cells were washed three times for 10 minutes, incubated with secondary antibodies, washed again, and mounted in Prolong Gold antifade containing DAPI (Invitrogen). Confocal microscopy was performed using an IX81 microscope (Olympus, Tokyo, Japan) with sequential excitation of fluorophores. The antibody against the mouse DIDO amino-terminal region (anti-DIDO NT) has been described before (Prieto, Kouznetsova et al. 2009). This antibody was raised against a HIS-tagged construct containing the mouse DIDO amino acids 2 to 90. All primary and secondary antibodies used for immunofluorescence experiments are depicted in **Table 6.**

| Primary antibodies | Dilution | Supplier company | Catalog number |
|---|---|---|---|
| anti-SFPQ | 1:200 | Abcam | ab177149 |
| anti-U2AF1 | 1:200 | Genetex | GTX106854, RRID:AB_1952473 |
| anti-hnRNP U | 1:200 | Abcam | ab20666, RRID:AB_732983 |
| anti-human DIDO (N-terminal) | 1:200 | Abcam | ab92868, RRID:AB_10562305 |
| anti-human NDUFS7 | 1:200 | ABGENT | AP20502C |
| anti-mouse DIDO (DIDO NT) | 1:100 | Lab production | - |
| **Secondary antibodies** | **Dilution** | **Supplier company** | **Catalog number** |
| Goat anti-Mouse IgG Cy3 | 1:400 | Jackson | 115-166-071 |
| Goat anti-Rabbit IgG Alexa 488 | 1:200 | Molecular Probes | A-11034 |

**Table 6**. Antibodies used in immunofluorescence experiments for human and mouse cell lines

To efficiently inhibit transcription elongation, RPE-1 cells were treated overnight with 500 µg/ml of Actinomycin D (A9415, Sigma, St. Louis, MO). To impair the splicing process and complex B formation, RPE-1 and MEF cells were treated overnight with 40µM of Isoginkgetin (416154, Merck Millipore). Both treatments were used as described elsewhere (Shav-Tal, Blechman et al. 2005, Huranová, Ivani et al. 2010). Besides, in both cases, vehicle cells were treated with DMSO.

The ImageJ package (rsbweb.nih.gov/ij) was used for image processing and measurements. For measurement of supramolecular complexes when MEFs were treated with IGK, contrast optimized images were binarized in ImageJ using watershed segmentation (Soille and Vincent 1990) before particle size analysis. Around 100 different cells were analysed to generate statistics. Representative images are shown in all figures.

## 6. Proximity ligation assay

Nuclear interaction between SFPQ and DIDO3 was detected using Duolink® *In Situ* Orange Starter Kit Mouse/Rabbit (DUO92102, Sigma), according to the manufacturer's instructions. The protocol was carried out as follows. MEFs were cultured on glass coverslips until 50% confluent. Cells were washed once in PBS, fixed with 3.7% FA (10 minutes), and washed again in PBS. Cells were permeabilized with 0.5% NP40 (5 minutes) and blocked using the provided blocking solution (1 hour, 37°C). After incubation with primary antibodies against DIDO and SFPQ (**Table 7**) for 1 hour at room temperature, two proximity ligation probes were added and incubated (1 hour, 37°C). DNA was amplified by adding polymerase and complementary fluorescent nucleotides (90 minutes, 37°C). Coverslips were mounted on Superfrost ultra plus microscope slides (Thermo Fisher Scientific, Waltham, MA) in the provided mounting medium, and visualized as described for immunofluorescence.

| Primary antibodies | Dilution | Supplier company | Catalog number |
|---|---|---|---|
| anti-SFPQ | 1:200 | Abcam | ab177149 |
| anti-mouse DIDO (DIDO NT) | 1:100 | Lab production | - |
| anti-RNAPII CTD (8WG16) | 1:200 | Abcam | ab817 |

**Table 7.** Primary antibodies used for Proximity Ligation Assay

To quantify the number of hotspots, 3D Object Counter plug-in from ImageJ (rsbweb.nih.gov/ij) was used. Around 20 cells were selected from each sample to measure the interactions. First, a low number of cytoplasmic dots were extracted from the picture. Once only the cell nuclei was selected, the 3D Object Counter plug-in quantified the total number of dots. In parallel, the area of every cell nucleus was measured for each counted cell. Finally, the number of dots were normalized to each cell area and statistical analysis were performed using a two-tailed Welch's t-test comparing WT with both *Dido* mutants.

## 7. Western Blot

Protein samples were obtained from WT, ΔE16 and ΔNT MEFs cells using a 1X SDS Sample Buffer (60 mM Tris-HCl pH 6.8, 10% Glycerol, 2% SDS, 50 mM DTT, 0.05% Bromophenol) supplemented with complete protease inhibitors (04693132001, Roche Applied Science). Proteins were resolved on 8% SDS-PAGE and transferred to a PVDF membrane. Antibodies used for *Dido* mutants detection are shown in **Table 8**. An antibody that recognizes specifically the mouse *Dido* exon 16 was generated in our lab and named as

anti-DIDO CT. This antibody was raised in rabbits against a peptide corresponding to the carboxy-terminal amino acids 1392 to 1410 of mouse DIDO3 (AFHTLLAEPGRPHDVQSVS).

| Primary antibodies | Dilution | Supplier company | Catalog number |
|---|---|---|---|
| anti-Tubulin | 1:2000 | Abcam | ab44928, RRID:AB_2241150 |
| anti-DIDO NT (N-terminal) | 1:500 | Lab production | - |
| anti-DIDO CT (C-terminal) | 1:500 | Lab production | - |

**Table 8.** Antibodies used for MEF ΔE16 verification.

## 8. RNA sequencing

WT, ΔE16 and ΔNT MEFs were grown to 80% confluence, washed once in PBS, and lifted by trypsinization. After washing again with PBS, total RNA was isolated from approximately $5{\cdot}10^6$ cells with a SPLIT RNA extraction kit (Lexogen, Austria) using the manufacturer´s protocol. For each mutant and the WT control, biological triplicates were prepared. Subsequent steps were outsourced to on-campus facilities (Parque Cientifico de Madrid). A Bioanalyzer (Agilent) was used to measure RNA quality and all samples used had RNA integrity numbers (RINs) of 9 or higher. Libraries were prepared from 1 µg total RNA by the TruSeq RNA v2 method (Illumina) and at least 50 million paired-end (2x100) reads per sample were obtained on a HiSeq 2000 system (Illumina).

## 9. RNA sequencing analysis

Sequencing adaptors were removed from both ends of the reads. Obtained sequences were aligned with Tophat2 (version 2.1.0) against the *Mus musculus* mm10 assembly (Daehwan Kim, Pertea et al. 2013) using the corresponding gene annotation model to guide alignment of splice junctions. Initial classification of splicing alterations was performed by Altanalyze software (Emig, Salomonis et al. 2010). The fragcount script from the RSeQC package (Huang, Li et al. 2012) was used to determine expression levels. To quantify exon skipping, the number of junctions spanning each exon was counted and divided by the corresponding expression (in FPKM) normalized to gene level. Some intergenic junctions were found, but these were not used for 3'SS or sequence motif identification. Using gene-level expression allows for quantitation of exons that are always skipped in one of the samples. Altered exons were identified as those that were significantly different ($p<0.05$) in a two-tailed t-test comparing triplicate *Dido* mutant samples to the

same number of WT controls. For 3′ SS usage, junctions mapping within 3 bases of known exon boundaries were analysed, and normalized according to the total number of junctions sharing the 5′ SS but not coinciding on the same 3′ SS. Altered usage was defined as a significant difference ($p<0.05$) of the ratio between the sampled and alternative 3′ SS, in a two-tailed Welch´s t-test comparing WT and mutant triplicates. Since the t-test is intrinsically conservative and our data showed a large degree of interdependence, no p-value correction was carried out.

## 10. Reverse-transcriptase PCR

Total RNA was isolated as described for RNA sequencing. Reverse-transcriptase PCR amplifications on 1 µg total RNA were performed using a Verso One-step RT-PCR Kit (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. To analyse splicing alterations, several set of primers were used to amplify specific regions of *Gas2L1*, *Fat1*, *Wdr8*, *Ndufs7* and *Dazap1* mouse transcripts. An additional set of primers was used to amplify the human *Gas2L1*, *Spata7*, *Ndufs7* and *Dazap1* genes. All primers are depicted in **Table 9 and 10**.

| Primers | Sequence (5' -3') |
|---|---|
| *mGas2L1 Long Isoform* | **Forward:** AAGCATGACCCTTGTCGCTG<br>**Reverse**: GAGGCTGAGGAGTCACTGTC |
| *mGas2L1 Short Isoform* | **Forward**: AAGCATGACCCTTGTCGCTG<br>**Reverse**: CAAACCCCAGAAGCCTGGAAA |
| *mFat1 Exon inclusion* | **Forward:** GTCAGAATGGAGGCGTTTGC<br>**Reverse**: GCAGAGGACGAACACCATCA |
| *mFat1 Exon skipping* | **Forward**: GGTGCTGTCTGTCAGTGTGA<br>**Reverse**: GCAGAGGACGAACACCATCA |
| *mWdr8 cUTR* | **Forward:** GCTGTTTGTGGTACTAGAGC<br>**Reverse**: AGACAGAAGTGGTCCTTACT |
| *mWdr8 aUTR* | **Forward**: GCTGTTTGTGGTACTAGAGC<br>**Reverse**: TGCCTTAAATGTGCTGTCAA |
| *mNdufs7 gene* | **Forward:** GCTATGTGGTGTCCATGGGG<br>**Reverse**: CACGCTTGATCTTCCGTTGC |
| *mDazap1 gene* | **Forward**: CGCCGACGAAATCGGGA<br>**Reverse**: TTTGCTGCTGTCGCTCCTG |
| *mNdufs7-Dazap1 Fusion* | **Forward**: GCTATGTGGTGTCCATGGGG<br>**Reverse**: TTTGCTGCTGTCGCTCCTG |
| *mNdufs7 minigene* | **Forward:** GGTTGTTCGTGGCTGTGACC<br>**Reverse**: TTCACGCTTGATCTTCCGTTG |
| *mDazap1 minigene* | **Forward**: GCCTGGACTGGAGTACCACT<br>**Reverse**: TCTAGGGTGTGTGGTCTGCT |
| *mNdufs7-Dazap1 Fusion* | **Forward**: GGTTGTTCGTGGCTGTGACC<br>**Reverse**: TCTAGGGTGTGTGGTCTGCT |

**Table 9**. Primers used for Reverse transcriptase PCR amplification of mouse genes.

| Primers | Sequence (5' -3') |
|---|---|
| *hGas2L1 Long Isoform* | **Forward:** GCTACTCCGGGGACAGTGAC<br>**Reverse**: ATGAGTGCTGCCCGTCTCG |
| *hGas2L1 Short Isoform* | **Forward**: TTCCAGCTCCTCCGATGAAG<br>**Reverse**: CGTGAGGGTTTACGGTCTGG |
| *hSpata7 gene* | **Forward:** GAGCAACCTCTGTCCTTCCC<br>**Reverse**: GCTTACTGGAACCGAGCAGT |
| *hNdufs7 gene* | **Forward:** ACGGAGGAGGCTACTACCAC<br>**Reverse**: CCGAGGGCAGGTTTATTGAC |
| *hDazap1 gene* | **Forward**: CAAGAGACTCTGCGCAGCTA<br>**Reverse**: CTGTTATCGCTCCTGGGTCC |
| *hNdufs7-Dazap1 Fusion* | **Forward**: ACGGAGGAGGCTACTACCAC<br>**Reverse**: CTGTTATCGCTCCTGGGTCC |

**Table 10**. Primers used for Reverse transcriptase PCR amplification of human genes.

## 11. Real Time quantitative PCR (RT-qPCR)

Total RNA was isolated as described for RNA sequencing with the SPLIT RNA extraction kit (Lexogen, Austria) and quantified using NanoDrop. The Transcription First Strand cDNA Synthesis Kit (04896866001, Roche Applied Science) was used to obtain the cDNA from 1 µg of isolated RNA in a BioRad Themocycler. RT-qPCR reactions were carried out in MicroAmp™ Optical 384-Well Reactions with Barcode (#4309849, Applied Biosystems). A master mix of primers (shown in **Table 11 and 12**) and *5X HOT FIREPol EvaGreen qPCR Mix Plus* (ROX) (Solis BioDyne) was added to the amplified cDNA. The equipment used for quantitative PCR was an ABI PRISM 7900 HT (Applied Biosystems). The conditions for the RT-qPCR reaction were: 5 min at 95ºC, 40 cycles of 15 seconds at 95ºC and 60 seconds at 60ºC. A final dissociation step was added.

The relative quantification of each gene was calculated using the Expression Suite Software (Thermo Fisher Scientific Inc.) and based on the $C_T$ values. These values were generated by using the fluorescence increase data and the number of cycles obtained from the equipment. The $C_T$ values were corrected using housekeeping genes to calculate a $\Delta$Ct value. Finally, this number was compared between WT and mutant samples ($\Delta\Delta$Ct) and the relative quantification (RQ) value was generated using the equation $2^{-(\Delta\Delta Ct)}$. This RQ number was represented in every graph applying a logarithmic scale.

| Primers | Sequence (5' -3') |
|---|---|
| *mGas2L1 Long Isoform* | **Forward:** AAGCATGACCCTTGTCGCTG<br>**Reverse**: GAGGCTGAGGAGTCACTGTC |
| *mGas2L1 Short Isoform* | **Forward**: AAGCATGACCCTTGTCGCTG<br>**Reverse**: CAAACCCCAGAAGCCTGGAAA |
| *mFat1 Exon inclusion* | **Forward:** GTCAGAATGGAGGCGTTTGC<br>**Reverse**: GCAGAGGACGAACACCATCA |
| *mFat1 Exon skipping* | **Forward**: CACTGGTCAGAGATGTCAGAGT<br>**Reverse**: GCAGAGGACGAACACCATCA |
| *mWdr8 cUTR* | **Forward:** GCTGTTTGTGGTACTAGAGC<br>**Reverse**: AGACAGAAGTGGTCCTTACT |
| *mWdr8 aUTR* | **Forward**: GCTGTTTGTGGTACTAGAGC<br>**Reverse**: TGCCTTAAATGTGCTGTCAA |
| *mNdufs7 gene* | **Forward:** GCTATGTGGTGTCCATGGGG<br>**Reverse**: CACGCTTGATCTTCCGTTGC |
| *mDazap1 gene* | **Forward**: CGCCGACGAAATCGGGA<br>**Reverse**: TTTGCTGCTGTCGCTCCTG |
| *mNdufs7-Dazap1 Fusion* | **Forward**: GCTATGTGGTGTCCATGGGG<br>**Reverse**: TTTGCTGCTGTCGCTCCTG |
| *mRPL32* | **Forward:** TTAAGCGAAACTGGCGGAAAC<br>**Reverse**: TTGTTGCTCCCATAACCGATG |
| *mbeta 2 microglobulin (β2M)* | **Forward**: GCCTGTATGCTATCCAGAAAACCCC<br>**Reverse**: TGAGGCGGGTGGAACTGTGT |
| *mTPT1* | **Forward**: CATCAGCCATGACGAGCTGTT<br>**Reverse**: CTCTGTTCTACTGACCATCTTGC |

**Table 11.** Primers used for RT-qPCR amplification of mouse genes. RPL32, β2M and TPT1 were housekeeping genes.

| Primers | Sequence (5' -3') |
|---|---|
| *hDido2* | **Forward:** ACAGGCACGTCAAGGACCTC<br>**Reverse**: AGAAAGCCGGCGCTTACC |
| *hDido3* | **Forward**: ACTCTTGCCCTTTGAGGGACCA<br>**Reverse**: TCGGGTCCGCTTTTCGTCCA |
| *hNdufs7 gene* | **Forward:** ACGGAGGAGGCTACTACCAC<br>**Reverse**: CCGAGGGCAGGTTTATTGAC |
| *hDazap1 gene* | **Forward**: CAAGAGACTCTGCGCAGCTA<br>**Reverse**: CTGTTATCGCTCCTGGGTCC |
| *hNdufs7-Dazap1 Fusion* | **Forward**: ACGGAGGAGGCTACTACCAC<br>**Reverse**: CTGTTATCGCTCCTGGGTCC |
| *hHPRT* | **Forward:** GGAAAGAATGTCTTGATTGTGGAAG<br>**Reverse**: AAGGAACCAGTCCGTCATATTAGG |
| *hbeta actin* | **Forward**: AGCGAGCATCCCCCAAAGTT<br>**Reverse**: GGGCACGAAGGCTCATCATT |
| *hGAPDH* | **Forward**: AGAAGGCTGGGGCTCATTTG<br>**Reverse**: AGGGGCCATCCACAGTCTTC |

**Table 12.** Primers used for RT-qPCR amplification of human genes. HPRT, β actin and GAPDH were housekeeping genes.

## 12. Photoactivatable Ribonucleoside - Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP)

MEF WT, ΔE16 and ΔNT samples were used for the PAR-CLIP experiments that were carried out as described (Hafner, Landthaler et al. 2010). Cells were cultured in 2 x 15 cm dishes until reaching 80-90% confluence and treated overnight with 100 μM 4-Thiouridine (4SU). After two PBS washes, samples were placed on ice and irradiated with 0,15 J/cm$^2$ UV light at 254 nm as previously described. Once mRNA-proteins were crosslinked, cells were harvested and lysed on Lysis Buffer (20 mM Tris-HCl pH 7.5, 500 mM LiDS, 0,5% LiDS, 1 mM EDTA and 5 mM DTT) and then homogenized using a 27G needle. At this step, an input was taken and the negative control was treated with RNAse A. To capture the Poly (A) mRNA-protein complexes, samples were incubated with oligo(dT)$_{25}$ magnetic beads (NE Biolabs) for 1 hour at 4°C with gentle rotation. Afterwards, these beads were washed with buffers containing decreasing concentrations of LiCl and LiDS. Then, mRNA-protein complexes were eluted by RNAse A treatment for 30 minutes at room temperature. With this step, mRNAs were eliminated and only the crosslinked proteins remained. Finally, SFPQ was detected by western blot using an anti-SFPQ antibody (RN106PW, MBL International) and the protein levels were measured using ImageJ software and "Gels" plug-in.

## 13. Individual - nucleotide resolution Crosslinking and Immunoprecipitation (iCLIP)

MEF WT, ΔE16 and ΔNT cells were cultured in 2 x 15 cm dishes. Here, the iCLIP protocol was performed as described (Ule, Jensen et al. 2003). After reaching 80-90% confluence, cells were washed with PBS, placed on ice, and irradiated with 0,15 J/cm$^2$ UV light at 254 nm. Then, cells were lysed using Lysis Buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0,5% NP40, 0,1% SDS, 0,5% sodium deoxycholate and 1 mM DTT) and partially digested with 1:250 RNAse I dilution for 3 minutes at 37°C and 500 rpm shaking. Once the reaction was stopped with RNAse inhibitors, an input sample was taken and the rest was immunoprecipitated using Protein A Dynabeads™ (Thermo Fisher Scientific Inc.). SFPQ was precipitated with a polyclonal antibody (RN106PW, MBL International, Woburn, MA) and an isotype was used as negative control. After 1 hour of 4°C incubation, beads were washed several times in buffers with decreasing salt concentration. Samples were digested with

Proteinase K and total RNA was isolated with the SPLIT RNA extraction kit (Lexogen, Austria) using the manufacturer´s protocol. A Bioanalyzer (Agilent) was used to measure RNA length and quality. Finally, RT-PCR amplifications were performed using a Verso One-step RT-PCR Kit (Thermo Fisher Scientific Inc.) with a pair of primers designed for *Gas2L1* gene detection (**Table 13**).

| Primers | Sequence (5' -3') |
|---|---|
| *mGas2L1* | **Forward:** TCACTGTCCCCAGAGTAGCG<br>**Reverse**: CCGCGCCAGTGATACACATC |

**Table 13.** Primers designed for Reverse Transcriptase PCR amplification of mouse *Gas2L1* iCLIP RNA.

A different version of the iCLIP technique was carried out in order to validate DIDO3 binding to the RNA. This modification involved the RNA labelling using an EZ-Link™ HPDP-biotin reagent (Thermo Fisher Scientific Inc.). Similar to the PAR-CLIP technique, cells were treated with 100 µM 4-Thiouridine (4SU), until reaching 80-90% confluence, and crosslinked with 0,15 J/cm$^2$ UV light at 254 nm. After cell lysis, inputs samples were taken and total RNA was labelled with 50µl of EZ-Link™ HPDP-biotin (4mM stock), that reacts to the thiol group contained in the 4SU. RNA-protein complexes were immunoprecipitated using Protein G Dynabeads™ (Thermo Fisher Scientific Inc.) with the anti-DIDO NT antibody (ab92868, abcam, RRID:AB_10562305) and Protein A Dynabeads™ (Thermo Fisher Scientific Inc.) with the anti-SFPQ antibody (ab177149, abcam). After 1 hour of 4°C incubation, beads were washed several times in buffers with decreasing salt concentration and RNA-protein complexes were eluted using 1X SDS Sample Buffer (60 mM Tris-HCl pH 6.8, 10% Glycerol, 2% SDS, 50 mM DTT, 0.05% Bromophenol). Finally, after resolving proteins on 8% SDS-PAGE, RNA-protein complexes were visualized using a fluorescent eFluor710 streptavidin conjugate. Input membranes were developed with anti-DIDO NT and anti-SFPQ antibodies.

# RESULTS

# Chapter 1. DIDO C-terminal interactions suggest a role in RNA metabolism

## 1.1. DIDO-SFPQ interaction is mediated through the DIDO3/exon 16

As mentioned in the introduction, GST pull down experiments showed that DIDO3 interacts with several splicing factors. The main interactor was SFPQ (PSF), a known SF associated with PPT binding (Patton, Porro et al. 1993). To test what DIDO isoform was involved in the SFPQ interaction, we co-immunoprecipitated the DIDO proteins after expression in HEK293T cells. V5 tag plasmids were used to overexpress the individual DIDO isoforms. Membranes developed with an anti-SFPQ antibody revealed that only DIDO3, the longest isoform, interacts with SFPQ (**Figure 11**).

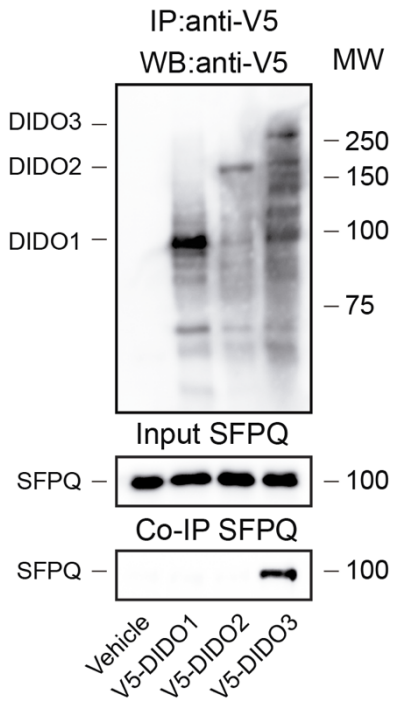



**Figure 11. DIDO-SFPQ interaction is specific for the DIDO3 isoform**. HEK293T cells were transfected with three plasmids bearing a V5 tag fused to mouse DIDO1, 2 or 3 isoforms. Even though the expression of V5-DIDO3 was lower than the other isoforms, and SFPQ levels are equal in all samples, WB revealed specific co-precipitation of SFPQ with DIDO3 longest isoform.

The results of the co-immunoprecipitation (Co-IP) supported the mass spectrometry analysis of the GST pull down, confirming the interaction between SFPQ and DIDO3. In addition, these data delimit SFPQ binding to the DIDO protein region corresponding to exon 16.

### 1.2. DIDO protein co-localizes with splicing factors when cells are treated with different RNA metabolism inhibitors

Most accessory splicing factors, such as SFPQ, have a function in positioning of spliceosome subunits on the nascent mRNA. However, only a small proportion of these co-localizes with other SF when splicing is proceeding normally. This is due to the dynamic behaviour of the spliceosome, where splicing factors are continuously recycled for their utilization in new complexes (Chen and Manley 2009, Kornblihtt, Schor et al. 2013). Small molecule inhibition of RNA metabolism has been widely used to interfere with this process. For example, Actinomycin D (Carmo-Fonseca, Pepperkok et al. 1991) acts by suppressing topoisomerases and transcription elongation (Sobell 1985). We thus evaluated the role of DIDO3 in the splicing process by studying SFPQ and DIDO distribution in the presence or absence of Actinomycin D (**Figure 12**). Confocal scanning laser microscopy was used to visualize the proteins.

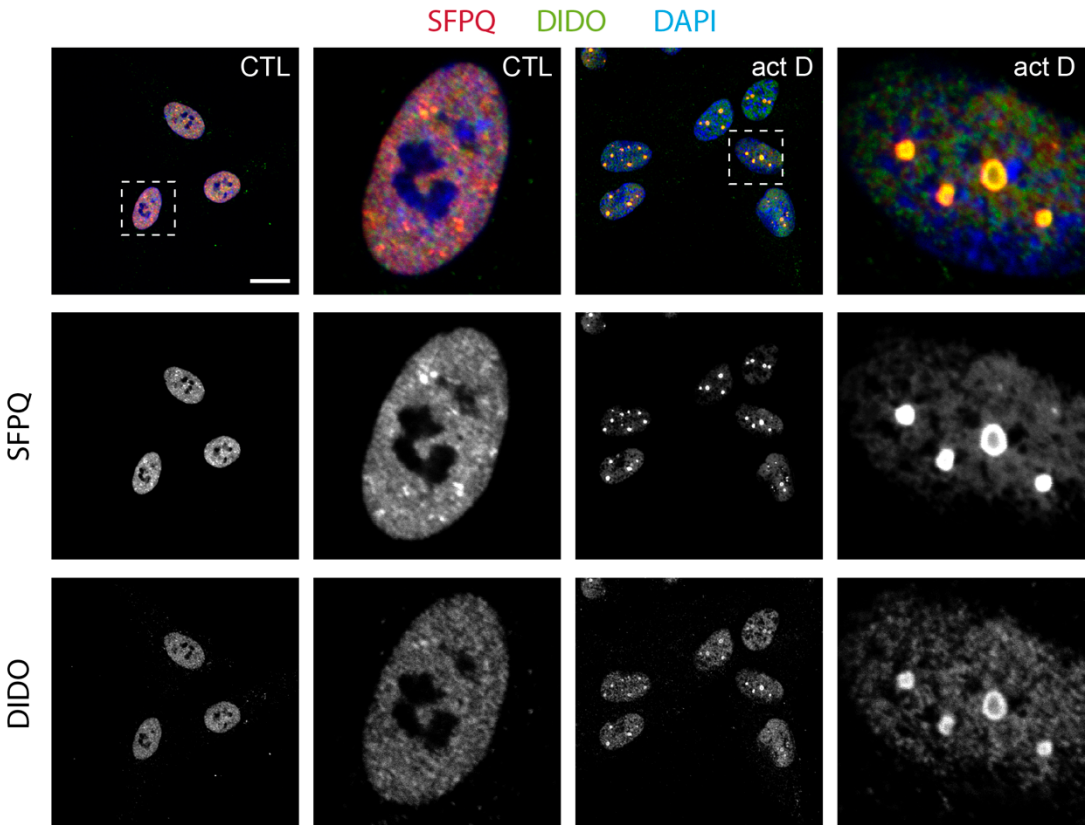
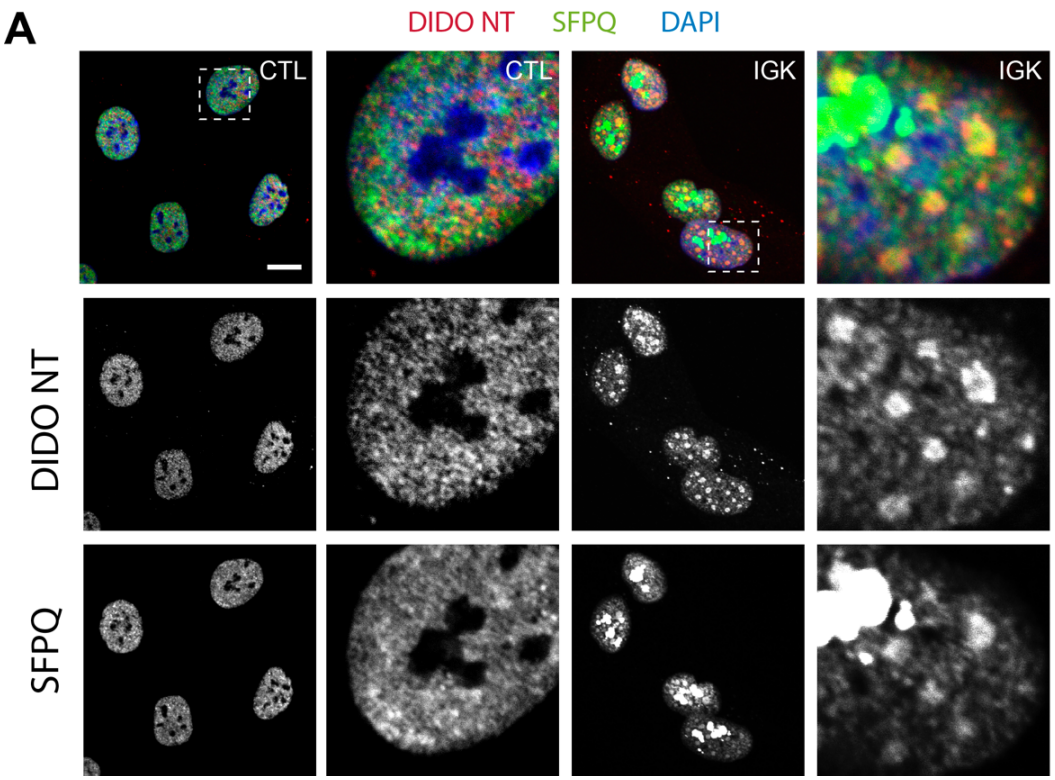



**Figure 12. Actinomycin D treatment produces co-localization of SFPQ and DIDO at subnuclear domains**. RPE-1 cells were seeded on coverslips and treated with vehicle (CTL) or 5 µg/ml of Actinomycin D overnight. First row shows the merge of three colours while second and third rows depict a colour separation of SFPQ and DIDO labelling respectively. The antibody against *H. sapiens* DIDO recognises the N-terminal part (DIDO NT), so it labels the three isoforms. Boxed areas show magnification of single nuclei. Scale bar, 10 µm.

While barely any co-localization was found in the control cells, DIDO and SFPQ accumulated at the same subnuclear domains after Actinomycin D treatment. This accumulation at nuclear speckles is typical of known splicing factors (Carmo-Fonseca, Pepperkok et al. 1991), which supports the idea that DIDO3 is involved in RNA metabolism.

Nevertheless, Actinomycin D can interfere with transcription and therefore with RNAPII recycling (Shandilya and Roberts 2012). We therefore used Isoginkgetin (IGK), an inhibitor that specifically affects the splicing process. IGK interferes with the U4/U6-U5 tri-snRNP recruitment to the spliceosome *in vitro*. As a consequence, the pre-spliceosome A accumulates and the U1 and U2 snRNP remain attached to the pre-mRNA, impairing their recycling (Behzadnia, Golas et al. 2007, O'Brien, Matlin et al. 2008, Huranová, Ivani et al. 2010). Not only SFPQ (**Figure 13A**) but also U2AF1 factor (**Figure 13B**) co-localized with DIDO in large subnuclear domains after IGK treatment. However, another splicing factor such as hnRNP U was not contained at these patches (**Figure 13C**).



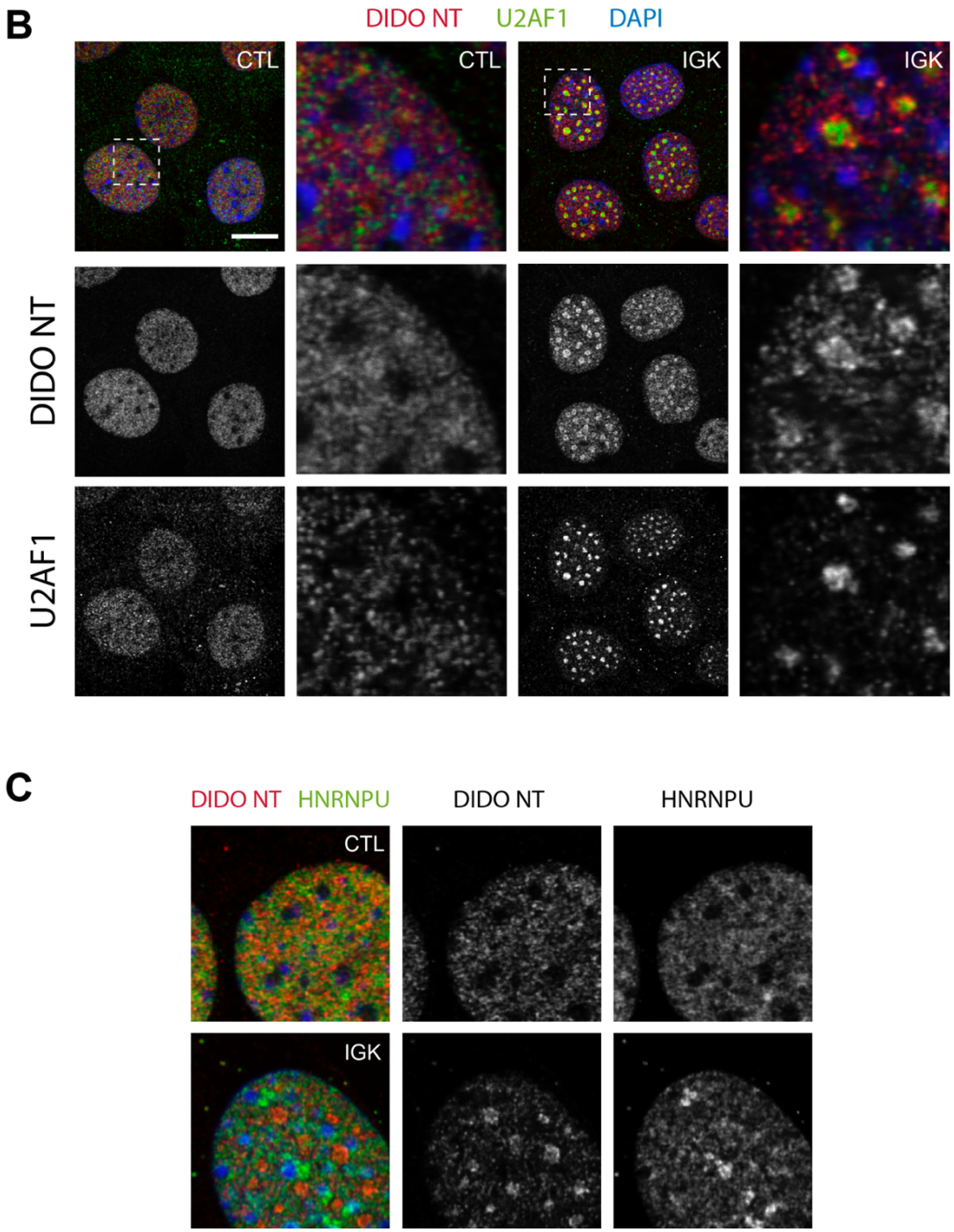



**Figure 13. Splicing inhibition produces the accumulation of some splicing factors and DIDO at subnuclear domains**. RPE-1 cells were seeded on coverslips and treated overnight with vehicle (CTL) or 40 µM of Isoginkgetin (IGK). (A) The first row shows the merge of three colours. The second and third rows depict a colour separation of DIDO NT and SFPQ labelling, respectively. (B) In this case, cells were labelled with U2AF1 and DIDO NT antibodies. (C) The first column shows the merge of three colours while second and third column depict DIDO and hnRNP U respectively. (A,B) Boxed areas show magnification of single nuclei. Scale bar, 10 µm.

These results show that DIDO, as well as SFPQ and U2AF1, is affected by splicing inhibition and changes its distribution in the cell nucleus by accumulating at typical

subnuclear domains. Unlike hnRNP U, both SFPQ and U2AF1 are involved in the correct positioning of the U2 snRNP on the 3'SS (Gozani, Patton et al. 1994). Based on protein localization, we hypothesize that the role of DIDO could be related with this step of spliceosome formation.

### 1.3. *Dido* mutants modulate IGK-dependent SFPQ distribution

Based on the DIDO3-SFPQ interaction and the IGK results, we asked whether this splicing inhibitor would also alter the SFPQ distribution in cells carrying *Dido* mutations. So, in order to test this hypothesis, Mouse Embryonic Fibroblast (MEF) cell lines lacking the entire exon 16 (ΔE16) or the N-terminal part (ΔNT) of the gene were used (**Figure 14**).

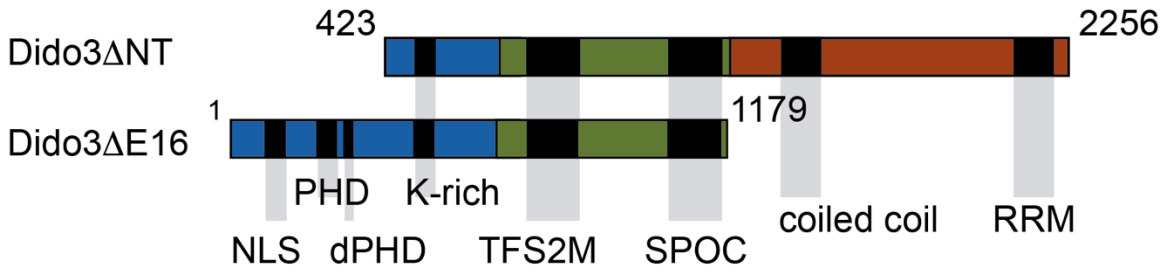



**Figure 14. Representation of *Dido* mutations generated in Mouse Embryonic Fibroblasts (MEFs).** Blue segments represent the region shared by the three DIDO isoforms, green depicts the central part of the protein, shared by DIDO2 and DIDO3 isoforms, and the carboxy-terminal part specific for the DIDO3 protein is shown in red colour. This C-terminal region corresponds to the *Dido* exon 16.

While *Dido* ΔNT mutants were made by homologous recombination (Futterer, Campanero et al. 2005), the *Dido* ΔE16 cells needed to be generated using a conditional strategy. Embryonic lethality was avoided by making the latter mutant with the Cre-LoxP system (**Figure 15A**). To obtain MEFs without the exon 16, a mouse strain bearing this exon flanked by *loxP* sites was interbred with Cre-ERT2 producing mice. The resulting embryonic fibroblasts were then isolated from the second-generation offspring. After 10 days of treatment with 4-hydroxytamoxifen, MEFs lacking *Dido* exon 16 were obtained and mutation was confirmed by PCR and WB experiments (**Figure 15B and 15C**).

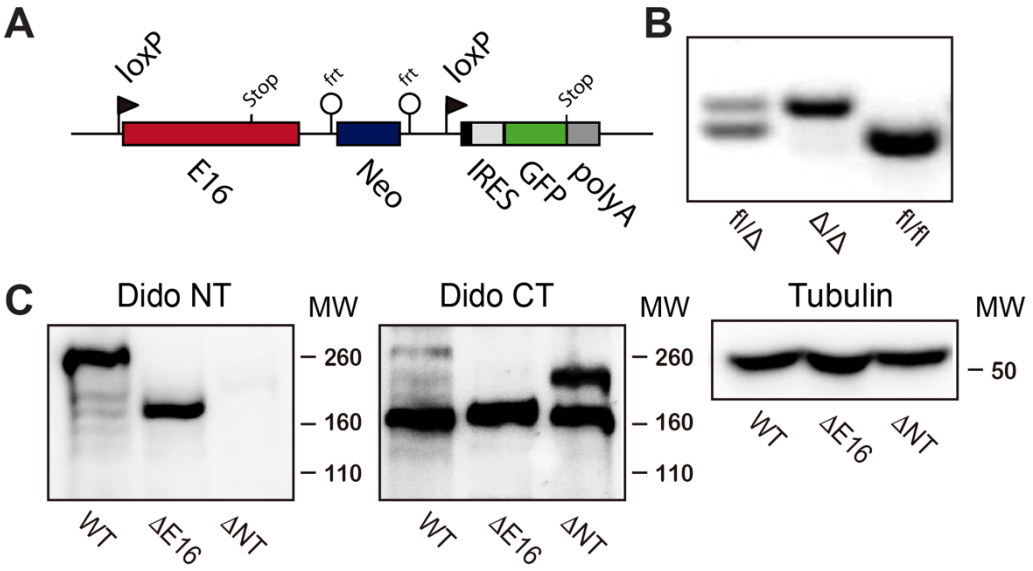



**Figure 15. Generation of MEFs lacking *Dido* exon 16.** (A) Schematic representation of exon 16 targeting and deletion strategy; (B) PCR confirmation of targeting (floxed/floxed; fl/fl) and deletion (Δ/Δ) of mouse *Dido* exon 16; (C) Western Blot confirmation of the deleted protein (ΔE16) using both DIDO antibodies, anti-DIDO NT that recognizes the N-terminal part of the protein and anti-DIDO CT that binds specifically to aminoacids encoded by the exon 16. Tubulin was used as a loading control and ΔNT mutant as a negative control for the DIDO NT antibody.

After obtaining both mutants, we decided to evaluate the effect of these *Dido* mutations on SFPQ distribution when splicing is affected. Thus, SFPQ location was studied by immunofluorescence after treating *Dido* WT, *Dido* ΔE16 and *Dido* ΔNT MEFs with IGK (**Figure 16**).

These analyses showed a speckled pattern corresponding to SFPQ distribution in MEF cells treated with IGK. This nuclear distribution is typical for SFPQ under conditions in which splicing is disrupted (Shav-Tal, Blechman et al. 2005) and already occurred in RPE cells (**Figure 13A**). However, whereas these subnuclear domains measured several square micrometers, *Dido* mutations seemed to reduce their size as compared to WT samples. In both mutants, especially in *Dido* ΔE16, IGK had little effect on SFPQ redistribution and most of the protein was located in smaller speckles (**Figure 16B**).

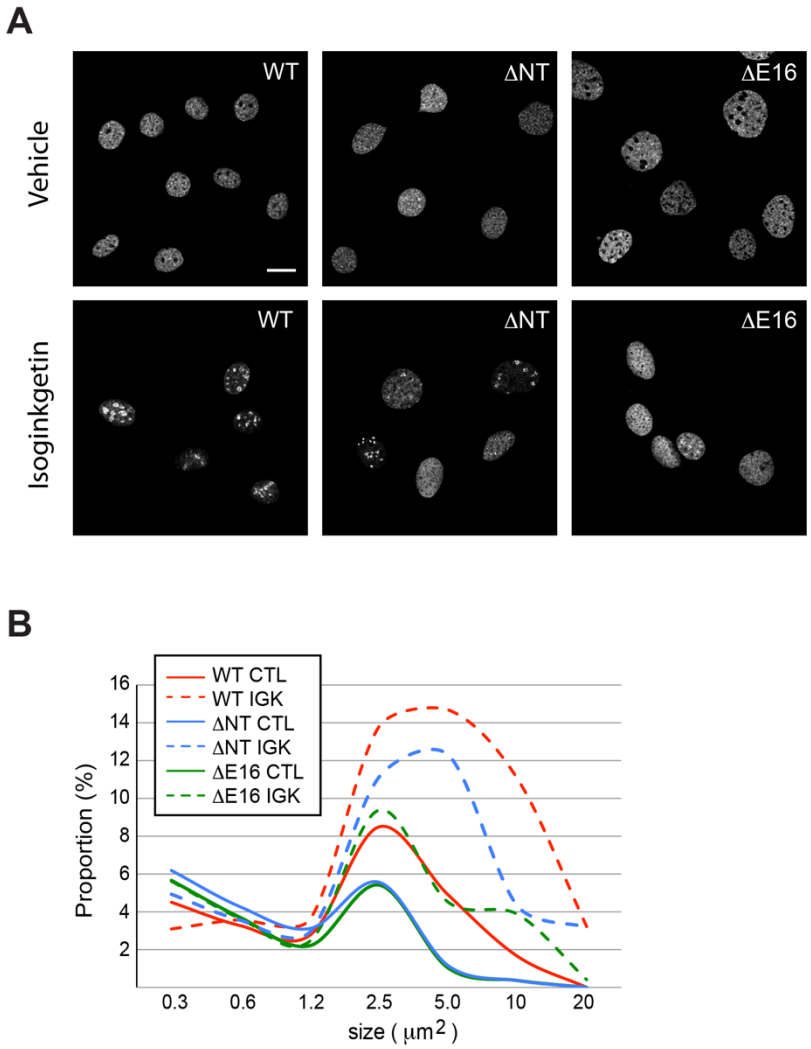



**Figure 16. *Dido* mutations reduce the size of supramolecular complexes after IGK treatment.** MEF WT, ΔE16 and ΔNT cells labelled with anti-SFPQ antibodies were treated with vehicle or 40 μM Isoginkgetin (IGK). (A) Representative immunofluorescence images of control (upper row) and treated cells (lower row). (B) Quantification of supramolecular complex sizes with and without IGK treatment in WT and mutant cells. The IGK treatment produced less supramolecular (>2.5 $\mu m^2$) structures in *Dido* mutants as compared to WT controls.

### 1.4. *Dido* mutation affects SFPQ recruitment

The previous data show that DIDO3/exon 16 controls SFPQ subnuclear distribution. To address whether *Dido* mutations could also alter SFPQ recruitment to the splicing machinery, Proximity Ligation Assays (PLA) were performed. In this technique, a quantifiable fluorescence signal is emitted when two proteins are interacting or at least closer than 40 nm (Weibrecht, Leuchowius et al. 2010). Thus, PLA experiments to quantify the DIDO3 and SFPQ interaction and detect possible recruitment problems were carried out with both *Dido* mutants. We included a first negative control with a single antibody to measure the background signal. In addition, the ΔNT mutant acted as an additional negative

control since the anti-DIDO NT antibody recognises the deleted region of this mutant. Both controls showed a low number of scattered PLA signals, confirming the validity of the technique. On the contrary, substantial interaction between both proteins was observed in WT MEFs (**Figure 17A**). Despite equal SFPQ expression in all samples (**Figure 17C**), the removal of exon 16 (*Dido* ΔE16) prevented PLA signal formation (**Figure 17A**). Further quantification of the PLA signals confirmed significant differences between WT and ΔE16 mutant MEFs (**Figure 17B**).

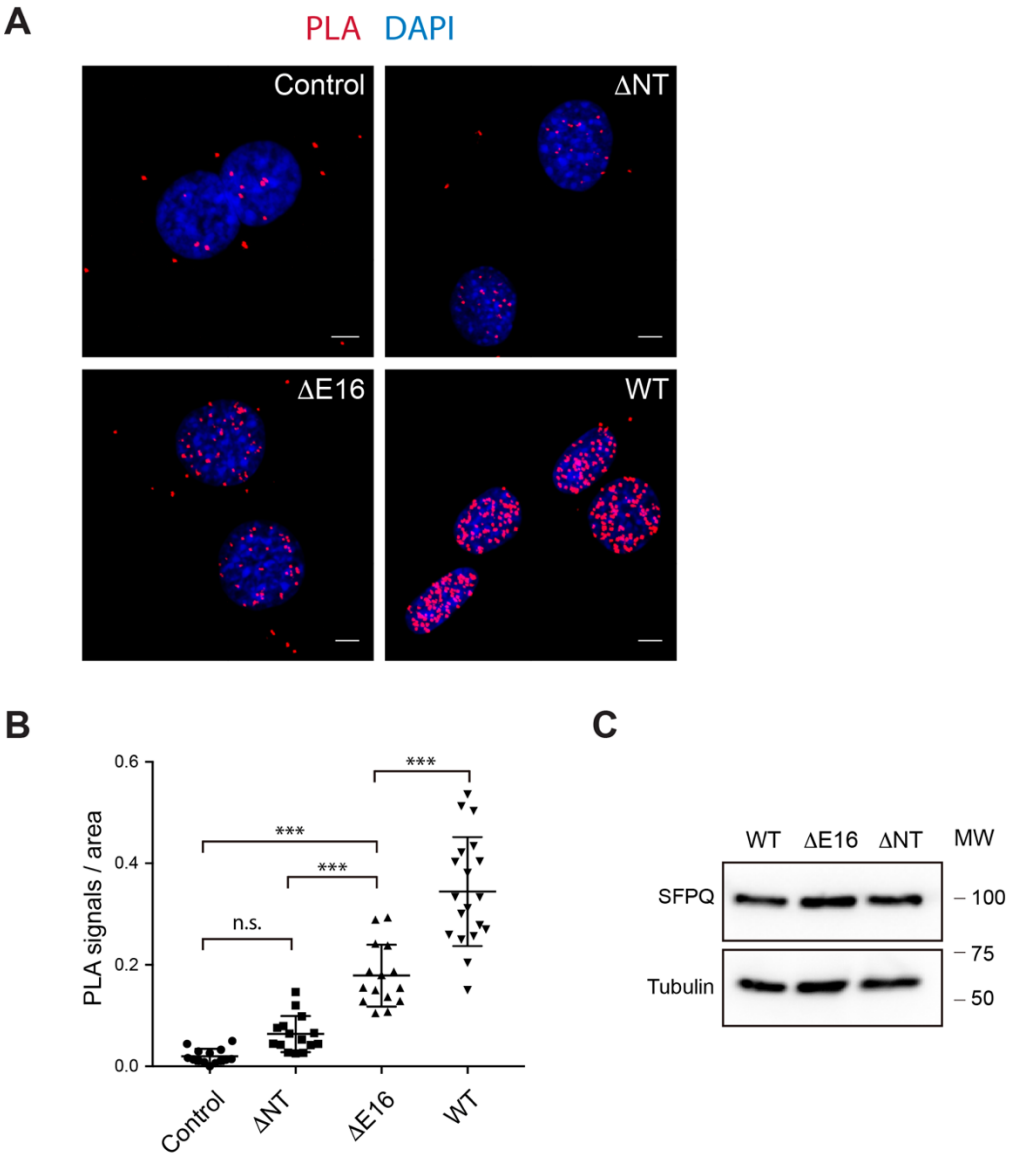



**Figure 17. DIDO3/exon 16 is necessary for SFPQ binding and recruitment**. (A) Immunofluorescence images of Proximity ligation assays between SFPQ and DIDO. Red spots represent proximity between both proteins, while nuclei are depicted in blue. The control sample was labelled with only one antibody, while ΔNT MEFs lacks the DIDO antibody epitope. ΔE16 MEFs lack the DIDO3 carboxy-terminal region that interacts with SFPQ. Proximity signals of DIDO and SFPQ were detected in WT but not in ΔE16 MEFs. (B) PLA signal quantitation revealed significant proximity loss in both mutants. Error bars represent standard deviation (SD). ***p<0.001. (C) WB of SFPQ protein levels in WT and mutant MEFs. Tubulin was used as a loading control.

Together, these results links the DIDO3 isoform to SFPQ recruitment and distribution at the cell nucleus. The distribution of DIDO3 is also affected by splicing inhibitors and follows SFPQ, implicating the protein in spliceosome formation at the U2 snRNP-positioning step. We therefore conclude at this point that DIDO3 is related to the splicing process, a hitherto undefined role for this protein. We thus designed experiments to further characterize this novel role, which are described in the next chapters.

### 1.5. DIDO as an RNA binding protein (RBP)

We explained in the introduction that DIDO3 contains a predicted RRM (RNA recognition motif) that is encoded within exon 16. Since similar domains are commonly found in RBP and splicing factors (Haynes and Iakoucheva 2006), DIDO3 not only could be implicated in the splicing process via SFPQ interaction but also through a direct binding to the RNA molecule. In addition, recent publications describing the RNA interactome included the DIDO protein in a long list of RBP candidates (Baltz, Munschauer et al. 2012, Castello, Fischer et al. 2012). In order to confirm DIDO3 binding to the nascent RNA molecule, we applied the PAR-CLIP technique.

PAR-CLIP is based on crosslinking by ultraviolet light and addition of a photoreactive uridine analogue named 4-thiouridine (4SU). The 4SU is inserted into nascent RNA during transcription. Subsequent UV irradiation at 254 nm generates a crosslink between RBPs and the RNA molecules. These mRNA-protein complexes are covalently bound, so pulling down the messenger RNAs using oligo (dT) beads allows us to isolate and analyse the associated RBPs (Castello, Horos et al. 2013).

To harness this technique for DIDO3, HEK293T cells were treated with 100 µM 4SU for 24 hours and subsequently UV irradiated. After cell lysis, the combination of messenger RNA and associated proteins was pulled down using oligo (dT) beads. Finally, through an RNAse treatment to remove all RNA molecules, RBPs were eluted and loaded on a protein gel to detect DIDO3 (**Figure 18A**). Nonirradiated cells incubated with 4SU were used as negative control. Despite seeing DIDO2 and DIDO3 in the inputs, we were not able to detect any protein in the crosslinked samples. The SFPQ control yielded a signal after crosslinking, showing effectiveness of the pull-down (**Figure 18B**).

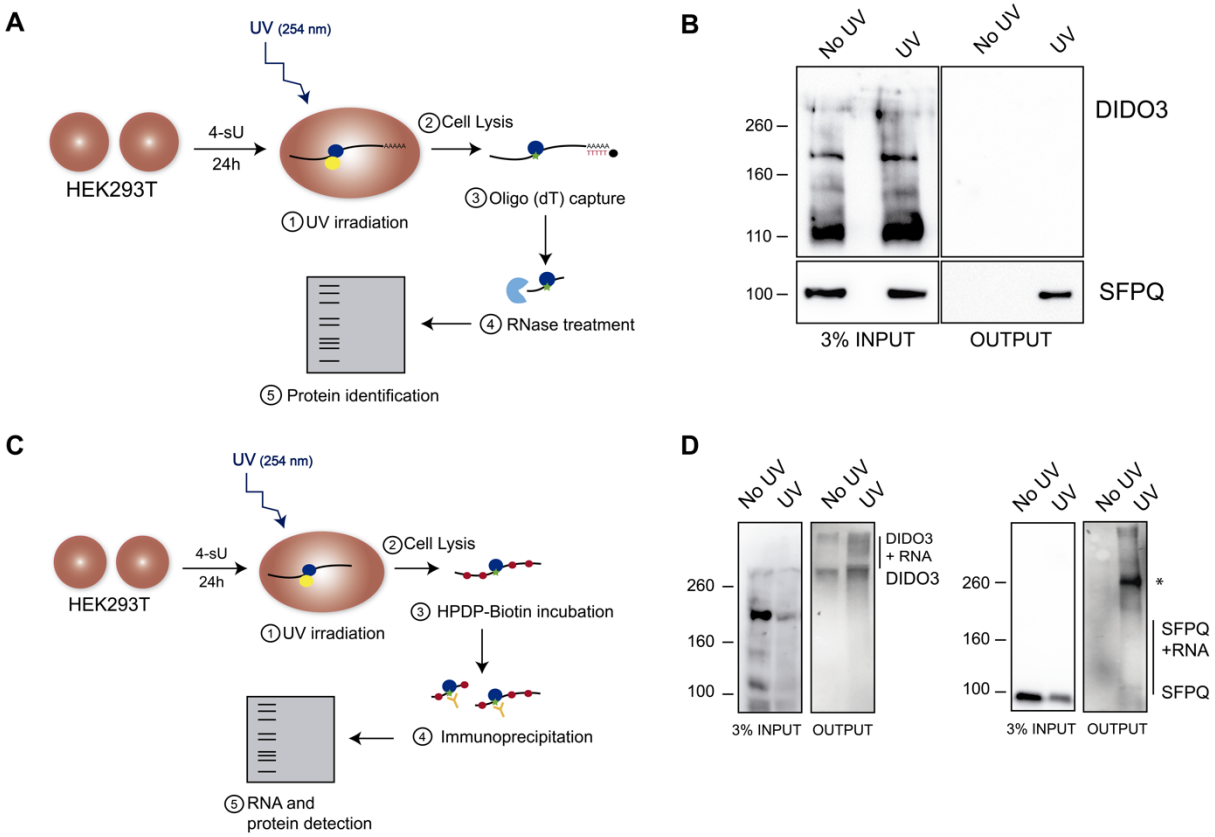



**Figure 18. iCLIP but not PAR-CLIP experiments suggest RNA binding by DIDO3.** (A) Schematic representation of the PAR-CLIP technique. HEK293T cells were grown until 50% confluence and treated for 24 hours with 100 µM 4SU. When they reached 90% confluence, cells were irradiated with 0,15 J/cm$^2$ UV light at 254 nm and lysed. After capture using oligo (dT) beads, the proteins were eluted with RNAse A, and DIDO and SFPQ were detected by SDS-PAGE. (B) Western Blot detection did not find co-precipitation of DIDO with messenger RNAs. Negative control samples were not UV irradiated and SFPQ detection was used as a positive control. (C) Schematic representation of the iCLIP technique. HEK293T cells were grown as described above. When they reached confluence, cells were irradiated with 0,15 J/cm$^2$ UV light at 254 nm and lysed. Then, RNA was marked with EZ-link HPDP-biotin. After immunoprecipitation, Protein-RNA complexes were separated by SDS-PAGE and RNA was detected with a fluorescent streptavidin conjugate. (D) Protein membranes show representative images of three different experiments. Output membranes were developed with an streptavidine-eFluor710 antibody to label the biotinylated RNA. Input membranes were developed with anti-DIDO NT (left) and anti-SFPQ (right) antibodies. Experiments correspond to the DIDO (left) and SFPQ (right) immunoprecipitations. Due to partial RNA degradation, a smear is observed above the protein band only in the UV irradiated samples. The smear is typical from RNA-protein complexes in both cases. An asterisk in the SFPQ output membrane indicates a possible SPFQ oligomerization or dimerization.

Since several proteins may compete for RNA in a PAR-CLIP, iCLIP was used to directly precipitate the target proteins (**Figure 18C**). In contrast to the PAR-CLIP experiment, iCLIP produced a typical smear in the crosslinked samples but not in the controls. This smear corresponds to partially degraded RNA covalently bound to the precipitated RBPs. The presence of a smear above DIDO3 thus indicates direct interaction with RNA (**Figure 18D**). Several factors, for example the proportion of interacting DIDO3 and protein stability, could

influence this alternative outcome; the epitope detected by the antibody used is located far from the predicted RRM and some protein degradation was evident in all DIDO samples. Both factors might have a greater impact on PAR-CLIP as compared to iCLIP.

Although one of these methods suggests DIDO3 binding to RNA, conclusive evidence is still missing. Thus, all subsequent experiments were based on the hypothesis that DIDO3 has a role on the splicing process through SPFQ interaction.

# Chapter 2. *Dido* mutations generate splicing defects

## 2.1. Different splicing alterations appear when *Dido* is mutated

To confirm the hypothesis that DIDO3 is implicated in RNA splicing, mutant MEFs lacking the N-terminal part (ΔNT) and exon 16 (ΔE16) were subjected to RNA sequencing. Triplicate samples of at least 50 million paired reads were analysed and compared to MEFs from WT littermates. Initially, we confirmed that the deleted region did not show residual RNA sequencing reads in the mutants (**Figure 19**).

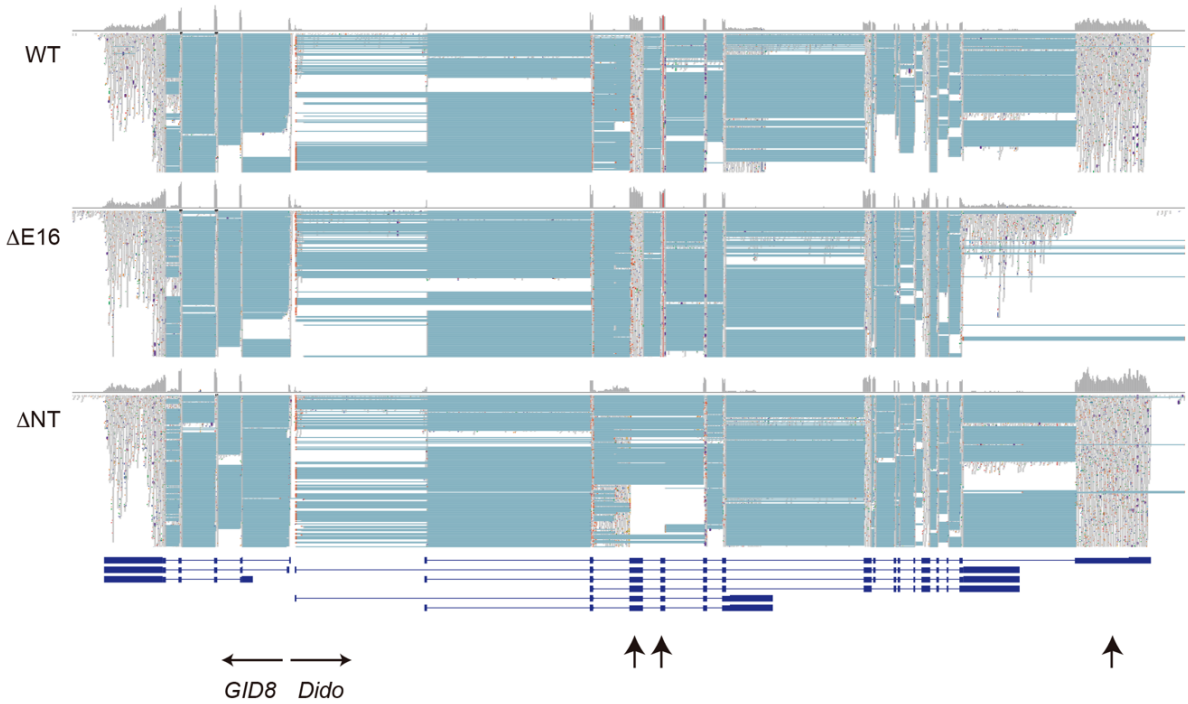



**Figure 19. RNA sequencing shows no residual reads in the *Dido* exons targeted for deletion** (arrows). Representative MEF clones are shown for WT and each mutant in the IGV viewer. Read density is depicted in grey and splicing events are connected by clear blue lines. Exons are represented as dark blue boxes and introns are shown as blue lines.

In order to identify splicing changes and knowing that several of the DIDO3-interacting proteins act in splice site selection, a first general classification of alternative splicing events was performed with the Altanalyze software (Emig, Salomonis et al. 2010). This analysis identified a large number (around several thousand) of splicing events that significantly differed between *Dido* mutants and WT controls (**Figure 20A**). The majority of these alterations involved exon skipping or exon inclusion, although a smaller subset corresponded to different events including alternative 5′ or 3′SS selection. Although most of

these changes have already been described in the literature or databases, a smaller percentage corresponds to novel events only visualized in our *Dido* mutants. When considering each mutant separately, ΔE16 has a preference for exon skipping events while the ΔNT mutant favoured exon inclusion.

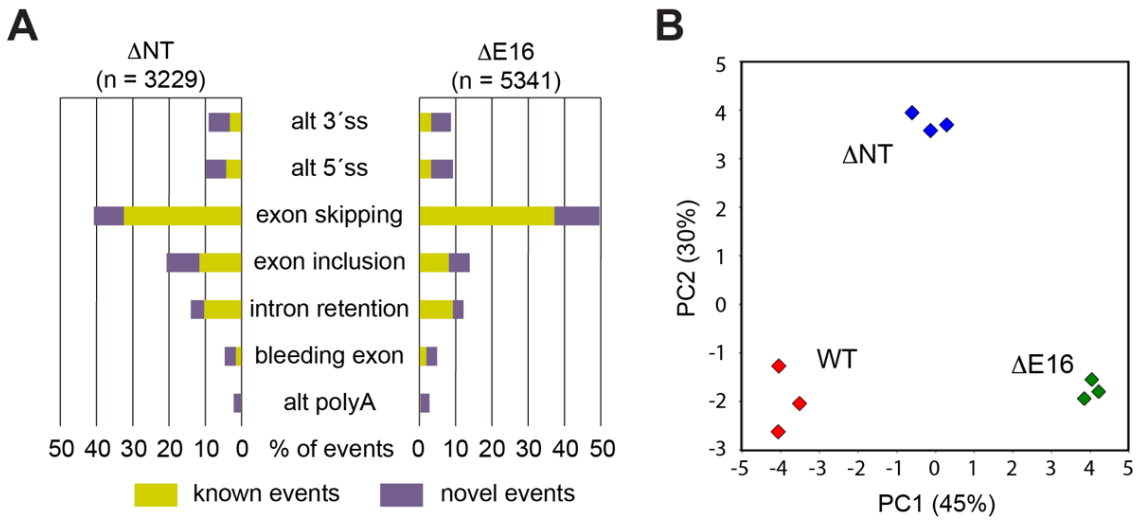



**Figure 20. *Dido* mutations generate several thousand of splicing alterations.** (A) General classification of the altered splicing events found in the mutants compared to the WT MEFs. Most of them (>60%) correspond to exon-level changes, such as exon skipping or exon inclusion. Yellow colour shows events already included in the databases, while purple boxes represent alterations not yet described. (B) Principle component analysis of altered splicing events distinguished WT, ΔNT, and ΔE16 groups.

Once we classified the alternative splicing events, a Principal Component Analysis (PCA) was performed to measure the relationship between the three sample groups. The results from this analysis produced three distinct groups, which corresponded to WT, ΔNT and ΔE16 samples (**Figure 20B**). Both mutants were displaced along one axis compared to WT controls (PC1), but only ΔNT moved up along the second axis (PC2). This separation indicates that both mutants share some qualities but at the same time produce unique alterations. Additionally, we quantified the levels of exon skipping with respect to WT controls in both mutants (**Figure 21A**). The volcano plots show that ΔNT mutants favour a situation of less skipping while ΔE16 samples bear widespread alterations encompassing exon skipping and inclusion. In the latter mutant, the most significant events are related with an increase in exon skipping. Finally, ΔNT and ΔE16 samples were compared in a Venn diagram depicting the significant events from the volcano plots (**Figure 21B**). With this comparison, we found that sizable exon populations suffer from the same alterations in both mutants but also uncovered many differences.

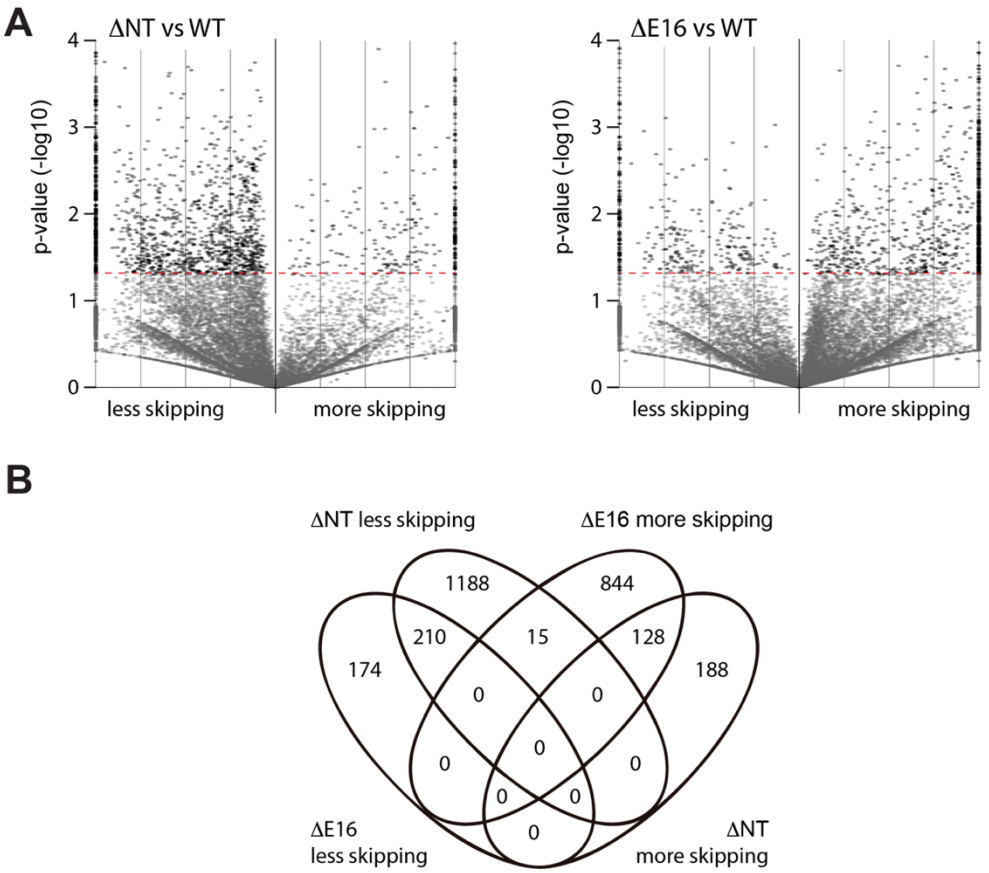



**Figure 21. Each mutant favours different types of splicing alterations.** (A) Volcano plots show exon skipping events in both mutants compared to WT. Increased skipping is depicted to the right of the vertical axis and decreased skipping to the left. The dotted line corresponds to the cut-off used for selection of significant events. (B) Venn diagram comparing exon skipping differences between mutants.

In order to confirm the RNA sequencing results, diverse subsets of exons were verified by real time quantitative PCR (RT-qPCR). In this thesis project, we will show four different examples that are representative of significant splicing alterations discovered by the massive sequencing (**Figure 22,23,24 and 25**).

The first example is the *Fat-1* gene. This member of the cadherin superfamily has a role in cellular polarization, in cell-cell contact, and in directing cell migration. In addition, it has been recently associated with the modulation of the epithelial-mesenchymal transition (EMT) and regulation of stemness genes in human cancer (Srivastava, Irshad et al. 2018). Based on our previous analysis, we located one internal exon that was commonly skipped in all samples. RNA sequencing showed that this exon was skipped significantly more in the ΔE16 mutant than in the WT sample (**Figure 22A**). Reverse transcriptase PCR and RT-qPCR confirmed the changes in exon skipping between samples (**Figure 22B and C**).

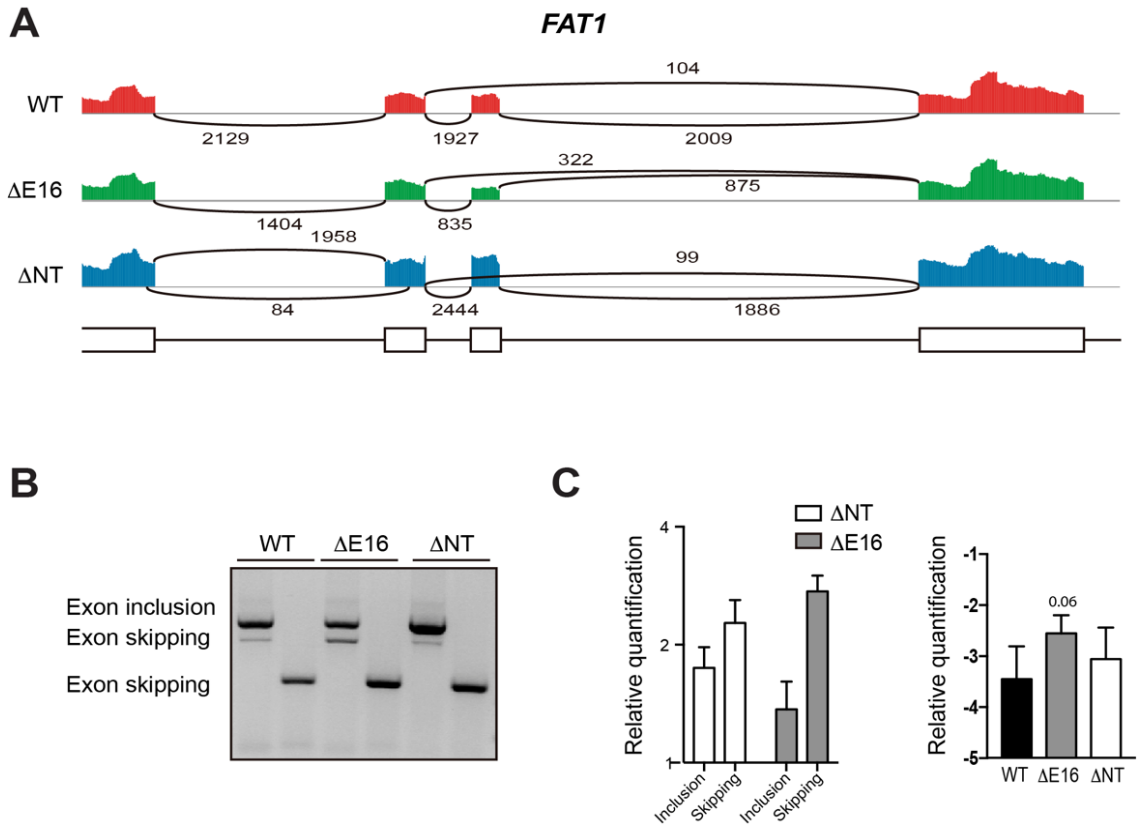



**Figure 22. DIDO3/exon 16 deletion increases exon skipping in the *Fat-1* gene.** (A) Sashimi plot of *Fat-1*. Expression coverage is shown in colour while splicing is represented by curved lines. White boxes denote exons and straight lines are introns. (B) Reverse transcription PCR experiments were performed to verify primer specificity. (C) RT-qPCR was used to quantify changes in exon skipping and inclusion. The first graph shows relative quantification comparing each mutant with WT samples. The second graph illustrates skipping differences between the three samples. Both graphs depict relative quantification values on Log 2 scale. Error bars represent SD from three different experiments.

A second example was found in the *Gas2L1* gene. This protein binds to microtubules and actin filaments, playing an essential role in centriole motility and centrosome disjunction (Au, Jia et al. 2017). GAS2L1 contains two different isoforms generated by alternative splicing. Although the WT sample produces the shortest transcript, a longer splice variant is generated when the last exon is skipped. This long transcript is highly enriched in both mutants while it is barely expressed in the WT sample (**Figure 23**). Interestingly, we found that the ΔE16 mutant favoured the expression of "downstream" transcripts in GAS2L1 and other similar examples. In all these cases, the constitutive and upstream exon was always skipped to include an alternative and downstream exon, usually generating different isoforms. Although this effect was also detected in ΔNT MEFs, differences were not as big as those found in the ΔE16 mutant.

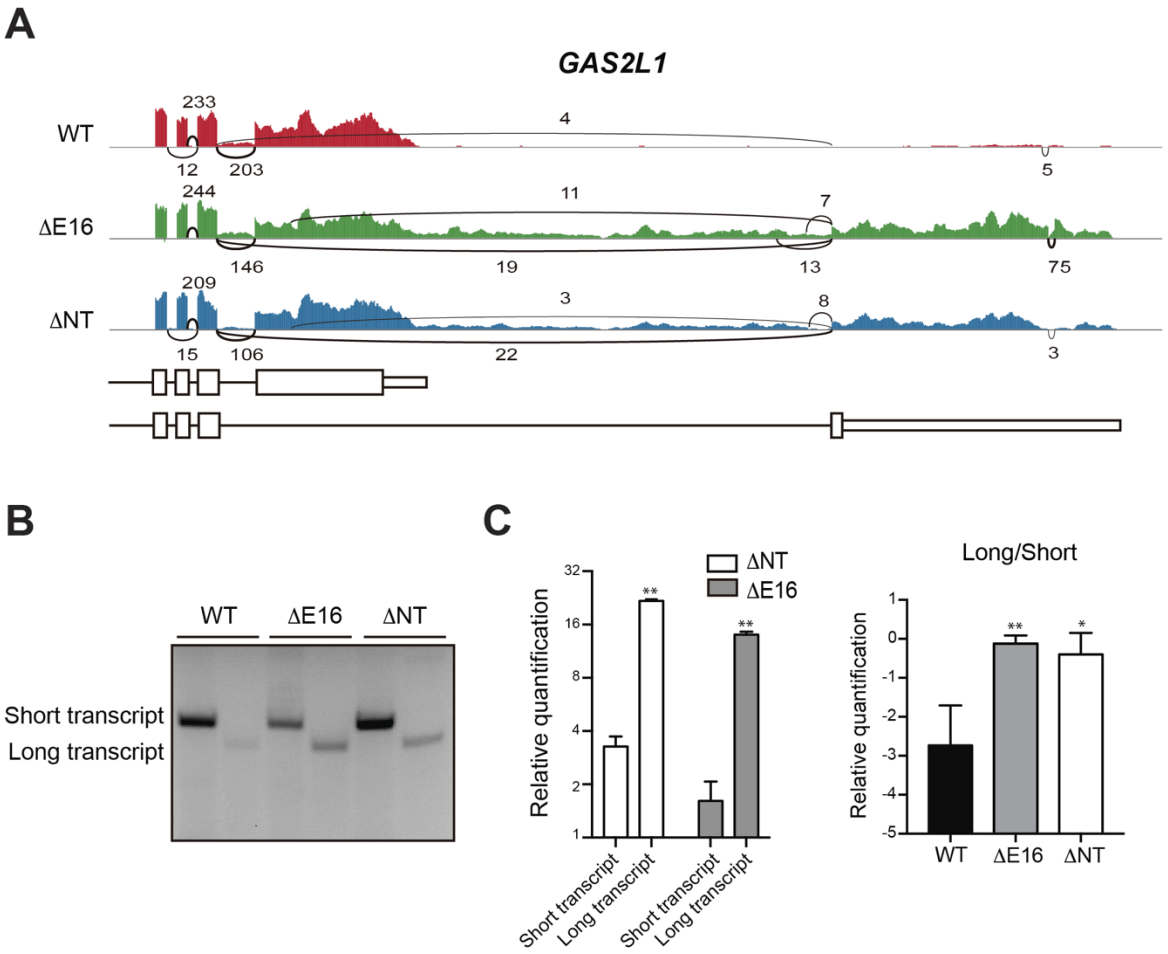



**Figure 23. *Dido* mutants favour last exon skipping in *Gas2L1* gene.** (A) Sashimi plot of *Gas2L1*. Expression coverage is shown in colour while splicing is represented as curved lines. White boxes denote exons and straight lines are introns. Upper scheme corresponds to the long isoform while the one below represents the short isoform. (B) Reverse transcription PCR experiments were performed to verify the primer specificity. (C) RT-qPCR was used to quantify both isoforms. The first graph shows relative quantification comparing both mutants with WT samples. The second graph illustrates the expression of the long transcript compared to the shorter one in all samples. Both graphs depict relative quantification values on Log 2 scale. Error bars represent SD from three different samples. *p<0.05; **p<0.01.

The *Wdr8* gene provided the next example. This WD repeat-containing protein acts as a regulator of spindle anchoring to the mitotic centrosome. A recent study described that WDR8 indeed is necessary for centrosome assembly in living vertebrate embryos (Inoue, Stemmer et al. 2017). The *Wdr8* gene contains a constitutive 3' UTR (cUTR) that generates the functional and complete WDR8 protein in WT controls. In both of our mutants, however splicing followed a different pattern. Here, splicing from the penultimate exon to an intergenic region downstream the constitutive 3'UTR was detected (**Figure 24**). This isolated region fits with the idea of an alternative 3'UTR (aUTR), as it contains some of the usual sequences for polyadenylation sites. In addition, NCBI databases report a non-coding RNA sequence that corresponds to this alternative UTR.

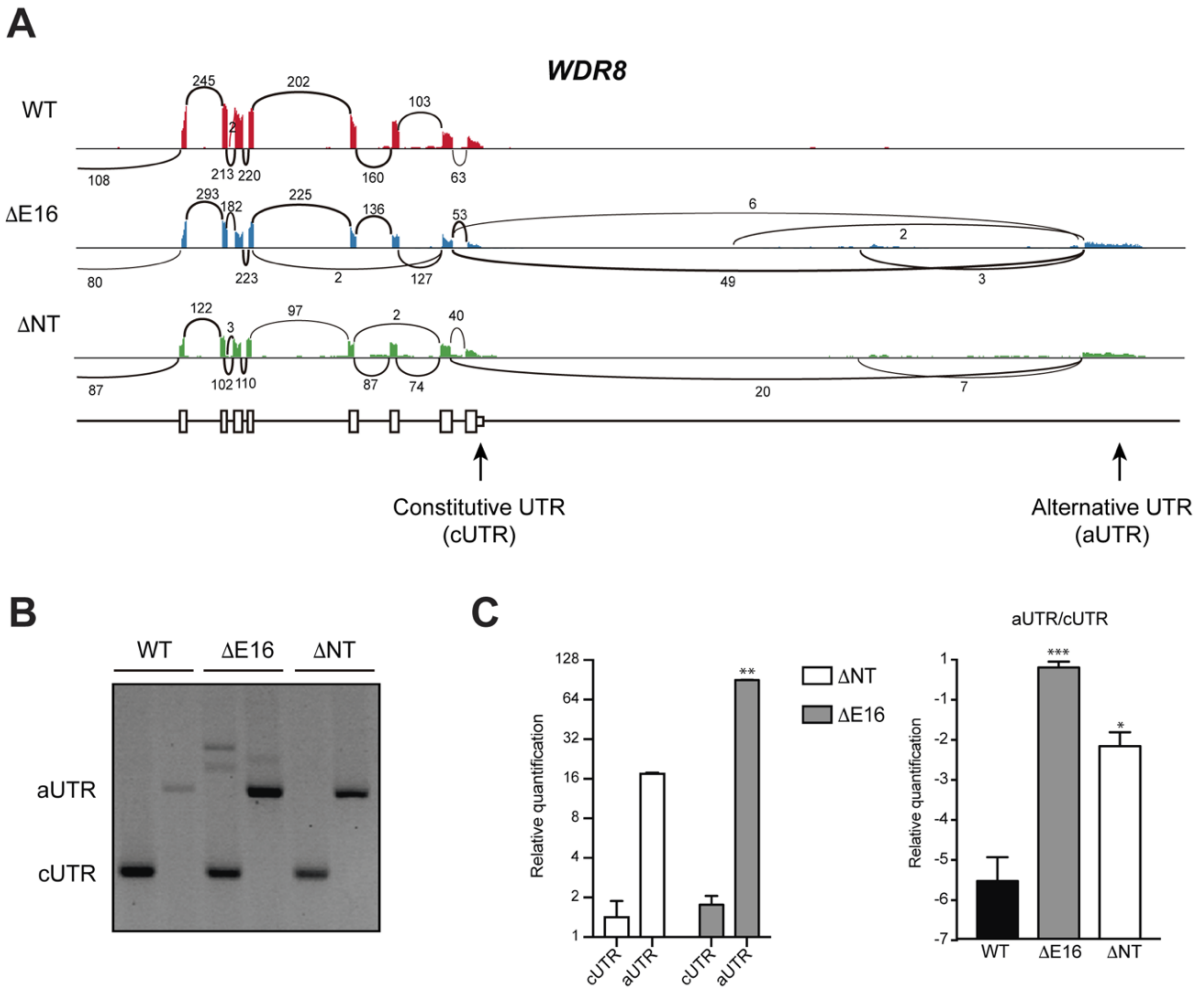



**Figure 24. *Dido* mutants promote the use of an alternative 3'UTR by the *Wdr8* gene.** (A) Sashimi plot of *Wdr8*. Expression coverage is shown in colour while splicing is represented as curved lines. White boxes denote exons and 3'UTRs, and straight lines are introns. (B) Reverse transcription PCR experiments were performed to verify primer specificity. (C) RT-qPCR was used to quantify mRNA containing the constitutive UTR (cUTR) or the alternative UTR (aUTR). The first graph shows relative quantification of both isoforms, comparing each mutant with WT samples. The second graph illustrates the expression of the aUTR compared to the cUTR in all samples. Both graphs depict relative quantification values on Log 2 scale. Error bars represent SD from three different samples. *p<0.05; **p<0.01; ***p<0.001.

Finally, our analyses identified a series of events that are rare under normal conditions, fusion messenger RNAs generated by transcription from two consecutive genes. In the last example, *Ndufs7* and *Dazap1* are two genes located in tandem on the same chromosome. NDUFS7 is a core subunit of the respiratory chain NADH dehydrogenase in the mitochondrial membrane (Hyslop, Duncan et al. 1996) and DAZAP1 is an RNA binding protein involved in mammalian development and spermatogenesis (Tsui, Dai et al. 2000).

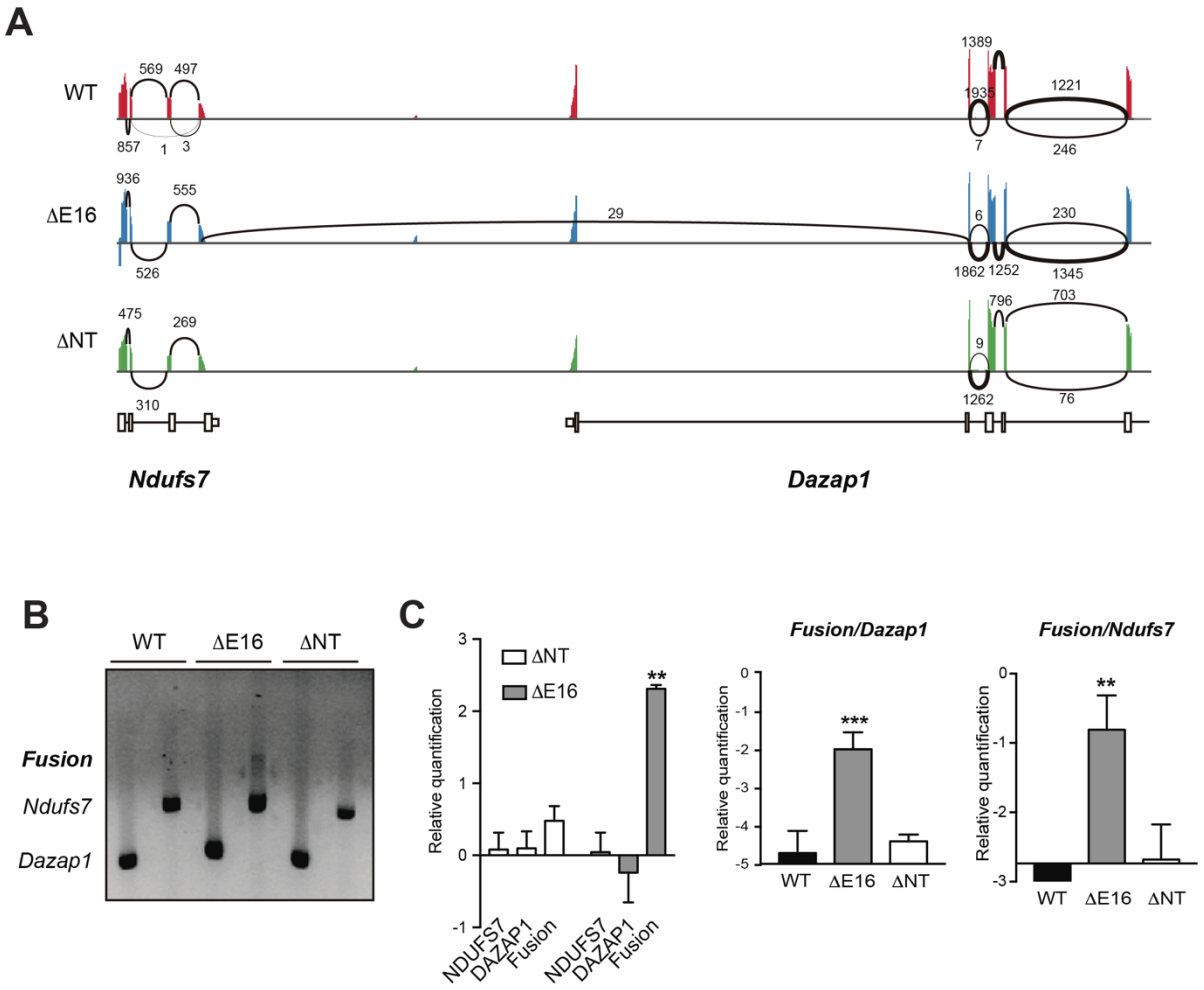



**Figure 25. *Dido* ΔE16 mutant generates a fusion messenger RNA between the consecutive genes *Ndufs7* and *Dazap1*.** (A) Sashimi plot of *Ndufs7* and *Dazap1*. Expression coverage is shown in colour while splicing is represented as curved lines. Splicing events between first and second *Dazap1* exons are not depicted in order to emphasise the intergenic splicing. White boxes denote exons and 3'UTRs, and straight lines are introns. (B) Reverse transcription PCR experiments showed amplification of the fusion mRNA. (C) RT-qPCR was used to quantify the mRNA from both genes and the fusion transcripts. In the first graph, we compared the level of each transcript in mutants to WT samples. The second and third graphs illustrate the amount of fusion relative to expression of each of the genes involved. The three graphs depict relative quantification values on Log 2 scale. Error bars represent SD from three different samples. **p<0.01; ***p<0.001.

We discovered that some splicing occurred from the penultimate exon of *Ndufs7* to the second exon of *Dazap1*, albeit only in the ΔE16 mutant. This aberrant splicing fused the messenger RNAs from both genes and generated a continuous transcript with shared regions. Using Reverse transcriptase PCR experiments, we confirmed the presence of this fusion mRNA. RT-qPCR quantified the frequency at which the fusion was produced in all samples (**Figure 25**).

## 2.2. Gene ontology analyses show an activation of the interferon-related response in *Dido mutants*

Once all splicing alterations were analysed, we asked whether gene expression was also affected and whether a specific pathway could be assigned to the effect of *Dido* mutations. Thus, Gene Ontology (GO) enrichment analyses and clustering were performed (**Figure 26**). GO enrichment showed altered expression of several pathways related to the interferon response. It seems that both *Dido* mutants display a coordinated regulation of interferon-stimulated genes, that has been associated with the cellular response to aberrant RNAs (Zevini, Olagnier et al. 2017).

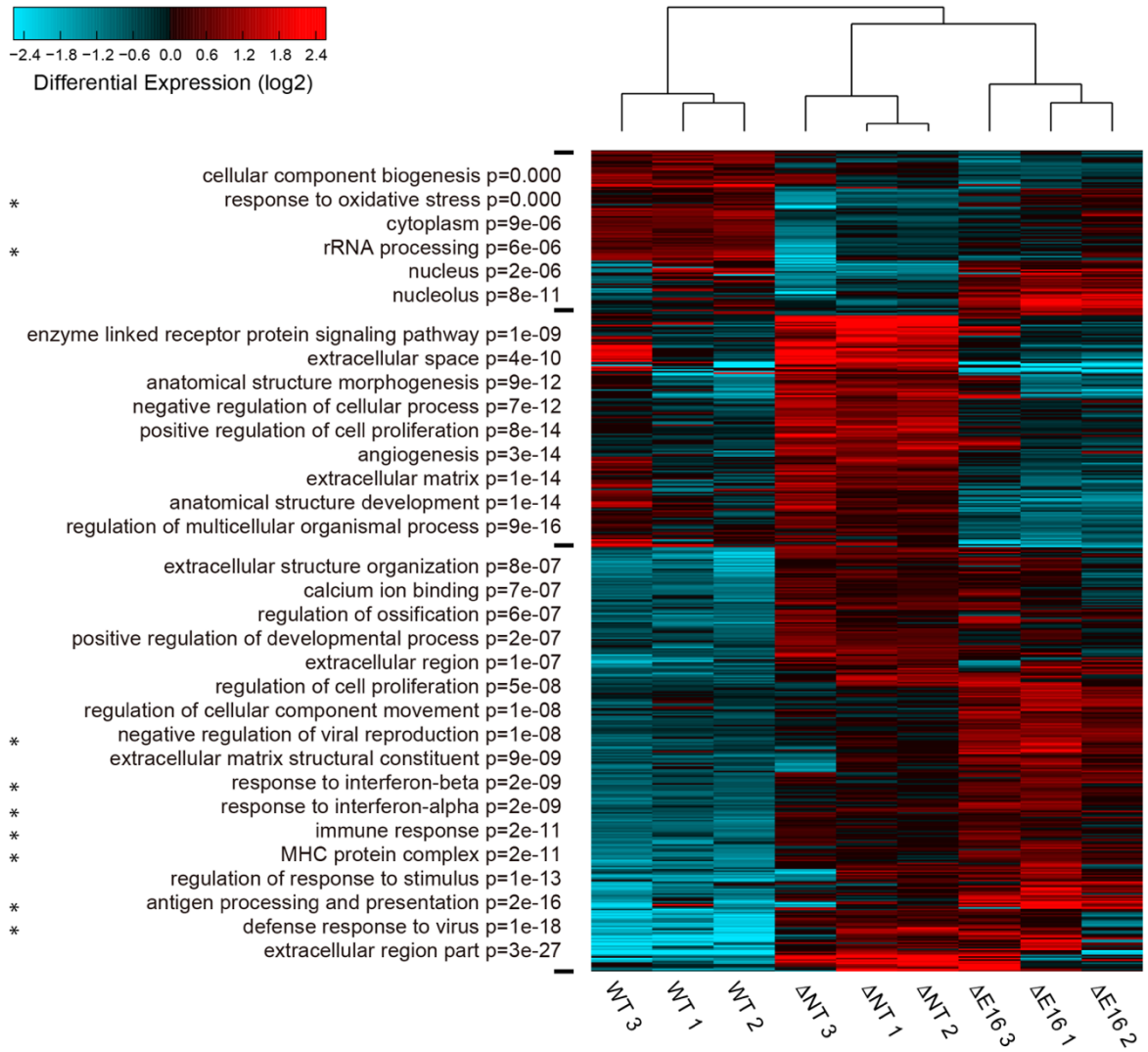



**Figure 26. *Dido* mutations produce an increase in the interferon-related responses.** GO analysis were performed using the Altanalyse program. Three replicates of each sample were included to obtain statistical power. The pathways showing significant enrichment or depletion were ordered according to the calculated p-value (left). The pathways associated with the interferon-related response were labelled with an asterisk.

## 2.3. Deletion of exon 16 in human cells produced splicing alterations similar to those found in mouse embryonic fibroblasts

In order to further study the role of DIDO3 in different contexts and bearing translational experiments in mind, we moved on to human cells. Since the majority of and the more varied splicing alterations were found in the ΔE16 mutant, we decided to generate a comparable mutation in a human cell line. We choose the human line HeLa Flp-In™ T-Rex™ (HeLa-FLP) stable cell line that contains a Tet-on system; these cells allow the introduction of a single copy of exogenous DNA into the HeLa genome, mimicking a relatively physiological situation for transgene expression. When desired, doxycycline supplementation induces the homogeneous expression from the introduced cassette.

In order to study the effect of DIDO3 deletion on these HeLa-FLP cells, we generated the corresponding mutation using the CRISPR/cas9 system. Since the deletion of exon 16 produced the strongest effects in mouse cells, we only reproduced this mutation. Design and cloning of the RNA guides were performed by Amaia Talavera, according to the described protocol (Ran, Hsu et al. 2013). The pSpCas9(BB)-2A-Puro vector, containing the Cas9 enzyme, was used for sgRNA cloning and transfection into the HeLa-FLP cell line. After selection with 2 µg/ml Puromycin for 48 hours, clones were genotyped to confirm correct exon 16 deletion (**Figure 27**).

Once mutant cells were obtained, we studied whether the same splicing alterations found in MEFs also occurred in the human cell lines. Thus, the *Gas2L1* gene already studied in mouse cells (**Figure 23**) and other susceptible candidates such as *Spata7* were selected. In the mouse *Spata7* gene, WT cells promoted inclusion of one of two mutually exclusive exons, while the ΔE16 mutants favoured inclusion of the other one (data not shown). We thus hypothesized that *Spata7* could suffer from the same kind of alterations in our mutant human cells. Indeed, the DIDO3/exon 16 deletion in humans favours exon skipping in the *Spata7* gene and the selection of the longest transcript in the *Gas2L1* gene (**Figure 28**). Similar to the situation in the mouse mutants, all skipped exons form part of non-canonical transcripts.

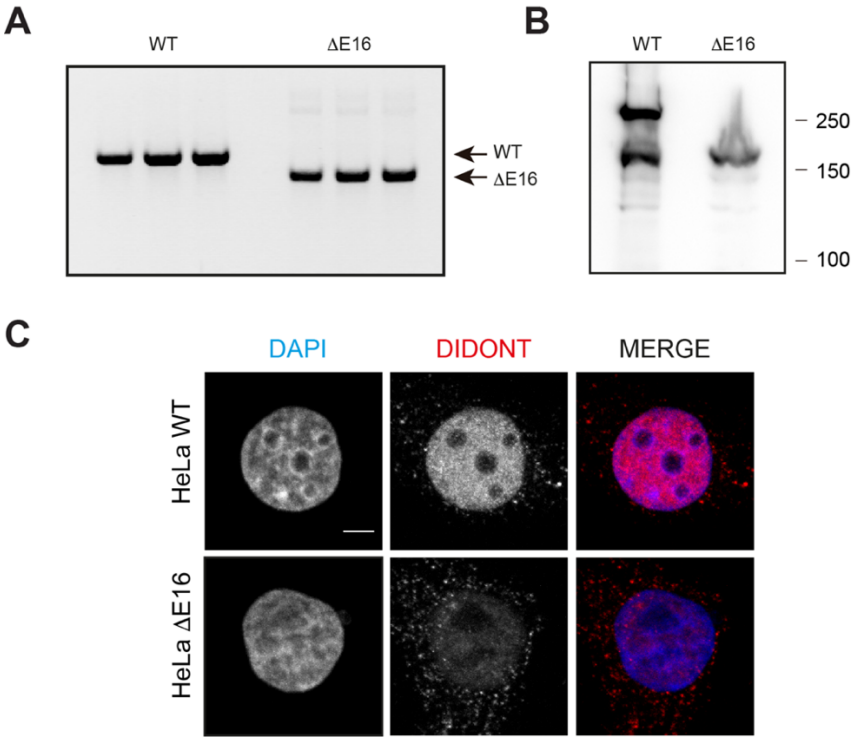



**Figure 27. Generation of HeLa-FLP cell lines lacking *Dido* exon 16 by CRISPR/Cas9.** We confirmed the elimination of the C-terminal region in several clones by PCR (A) and Western Blot (B) techniques. To validate absence of DIDO3, we used the anti-DIDO NT antibody that recognises the N-terminal part of the protein. Three different clones are depicted in (A) while a representative image of one clone is shown in (B). (C) Immunofluorescence analysis of HeLa WT and ΔE16 cells labelled with DAPI and anti-DIDO NT antibody. Scale bar, 5 μm.

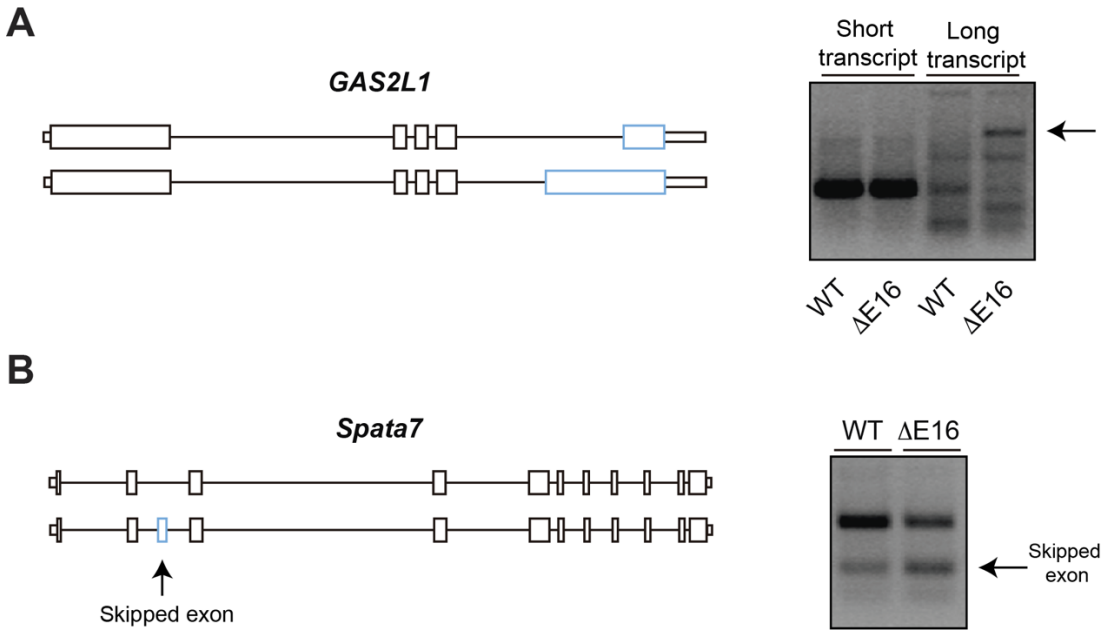



**Figure 28. Deletion of *Dido* exon 16 in human cells reproduces splicing alterations.** (A) The *Gas2L1* gene in human cells presents two isoforms generated by alternative splicing. While the short transcript is equally expressed in WT and ΔE16 samples, the long transcript can only be found in the mutant . The blue boxes represent the exons implicated in alternative splicing. (B) *Spata7* exon skipping events are increased in the ΔE16 mutants compared to the WT cells. The blue box depicts the skipped exon.

Another example is the fusion transcript generated between the consecutive genes *Ndufs7* and *Dazap1*. In our ΔE16 mouse mutants, splicing from the last exon of *Ndufs7* to the second exon of *Dazap1* was found. These alterations produced a fusion transcript containing regions of both genes. We therefore asked whether our human mutants would trigger the same fusions. Surprisingly, a fusion transcript was generated but splicing started in an upstream region as compared to the fusion in the mouse samples. Here, splicing was initiated from the penultimate exon instead of the last exon of *Ndusf7.* And although little signal was found in the WT samples, it seemed that more fusion transcript was generated by the ΔE16 cells (**Figure 29**).

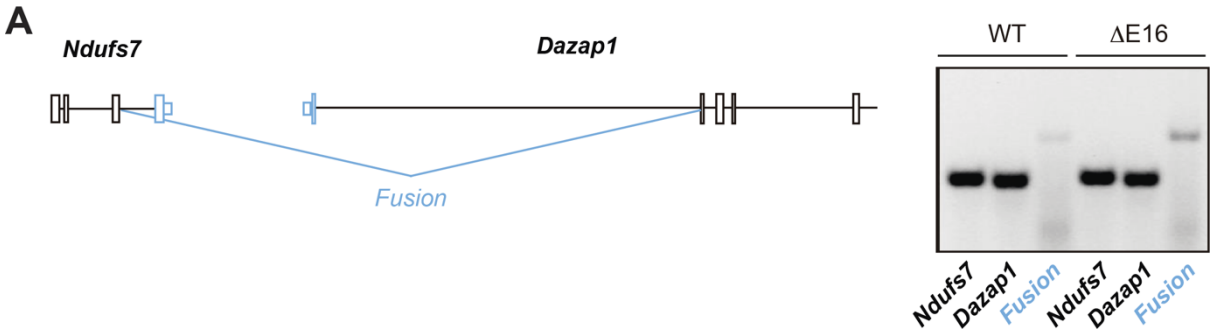



**Figure 29. Fusion transcripts are also produced in HeLa-FLP ΔE16 mutant cells.** The blue boxes represent the skipped exons and the blue line depicts the splicing event that generates the fusion transcript. Expression levels of *Ndufs7* and *Dazap1* are equal in both samples while the fusion transcript seems to be generated more frequently in the ΔE16 mutant.

Interestingly, the splicing events in the human mutants gave rise to an in frame fusion of *Ndufs7* and *Dazap1* exons, so the fusion transcript could be translated into a fusion protein. In order to test the generation of this protein, western blot and immunofluorescence analyses were performed (**Figure 30**). As expected, only the ΔE16 cells produced a fusion protein that was detectable with an anti-NDUFS7 antibody (**Figure 30A**). In addition, immunofluorescence experiments showed an antibody signal that was not exclusively located in the cytoplasm, although NDUFS7 is considered a cytoplasmic protein. In the exon 16 mutants, NDUFS7 signal was distributed more between the cytoplasm and the nucleus (**Figure 30B**). Since DAZAP1 is a nuclear protein, we believe that the fusion protein was targeted to the nucleus instead of being exported to the mitochondria.

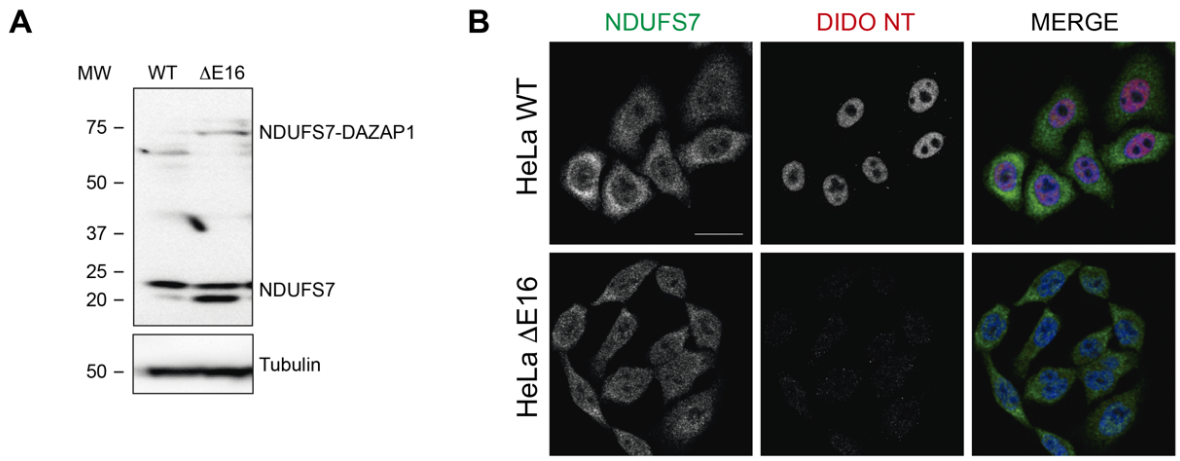



**Figure 30. The *Ndufs7-Dazap1* fusion transcript produces a fusion protein with shared properties.** (A) Western blot of WT and ΔE16 HeLa-FLP cells developed with anti-NDUFS7 (upper box) and anti-tubulin (lower box). We detected the NDUFS7 signal at around 20 kDa and an additional band at approximately 70 kDa only in the ΔE16 mutant. This upper band could correspond to the fusion protein formed by NDUFS7 (20 kDa) and DAZAP1 (50 kDa). Tubulin was used as a loading control. (B) Immunofluorescence experiments labelling HeLa-FLP samples with anti-NDUFS7 (green), anti-DIDO NT (red) and DAPI (blue). The NDUFS7 signal was mostly cytoplasmic in the WT cells but distributed between the nucleus and the cytoplasm in the ΔE16 mutants. The absence of DIDO3 signal in the C-terminal mutant confirmed exon 16 deletion. Scale bar, 25 μm.

These data show that the lack of exon 16 in human cells results in alternative splicing defects similar to those found in MEFs. In some cases, such as the *Ndufs7* and *Dazap1* intergenic splicing, not only a fusion transcript but also a fusion protein is produced. Moreover, this fusion protein is even capable of modifying the normal location of the NDUFS7 protein, implying that the splicing alterations can have a global effect.

## 2.4. Complementation with full-length DIDO3 restores the normal splicing metabolism in human cells

Since human *Dido* mutants showed splicing alterations comparable to mouse cells, we tested if complementation experiments with the full-length protein would revert this effect. As mentioned before, HeLa-FLP cells allow stable integration of exogenous DNA and induction of its expression by doxycycline. Thus, the full-length human *Dido3* cDNA *(hDido3)* was cloned in a pcDNA$^{TM}$5-FRT-TO-GFP vector and integrated in ΔE16 cells. Vice versa, we wanted to study whether saturation of the RNA polymerase II interaction through the TFS2M domain could affect RNA metabolism and therefore the splicing process. Thus, a complementary plasmid containing the human *Dido2* cDNA (*hDido2)* was generated, to overexpress this isoform in WT cells.

After obtaining clones, the expression of both proteins (DIDO2 and DIDO3) was induced with 1 µg/ml Doxycycline for 72 hours. Immunofluorescence analyses were used to detect GFP fused proteins in HeLa-FLP WT and ΔE16 cells (**Figure 31A**). Although the expression was not entirely homogeneous, most cells expressed the full-length DIDO3 isoform. In the case of DIDO2, the expression was lower than DIDO3 even though the same induction time was applied. However, the real expression levels measured by RT-qPCR do detect the overproduction of both human DIDO2 and DIDO3 isoforms in complemented cells (**Figure 31B**).

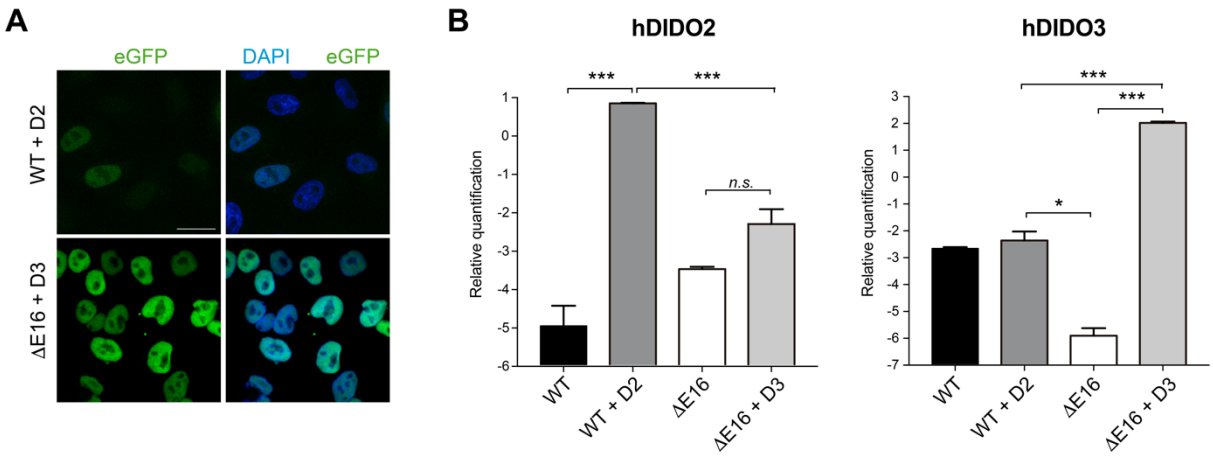



**Figure 31. hDIDO2 and hDIDO3 overexpression in human HeLa-FLP cells.** (A) HeLa-FLP cells were transfected with plasmids containing eGFP fused to human DIDO2 or DIDO3. Immunofluorescence experiments were performed to evaluate expression of eGFP fusion constructs. Although widespread expression of exogenous *hDido3* was found in the ΔE16 mutant, hDIDO2 induction was not as intense as expected. Scale bar, 25 µm. (B) Quantification by RT-qPCR of hDIDO2 and hDIDO3 expression levels in all samples. Graphs depict relative quantification values on Log 2 scale. In both cases, the primer pair not only detects the endogenous gene but also the inserted one. Error bars represent SD from three different samples. *p<0.05; ***p<0.001.

To test whether the splicing alterations were reverted by stable DIDO3 expression in our ΔE16 cells, we measure the levels of the *Ndufs7-Dazap1* fusion transcript. After correcting for expression of the endogenous genes, expression of the fusion transcript was significantly reduced in exon 16 mutants complemented with DIDO3. Interestingly, overexpression of DIDO2 increased fusion transcript levels compared to WT controls, although this effect was mild as compared to exon 16 absence (**Figure 32**).

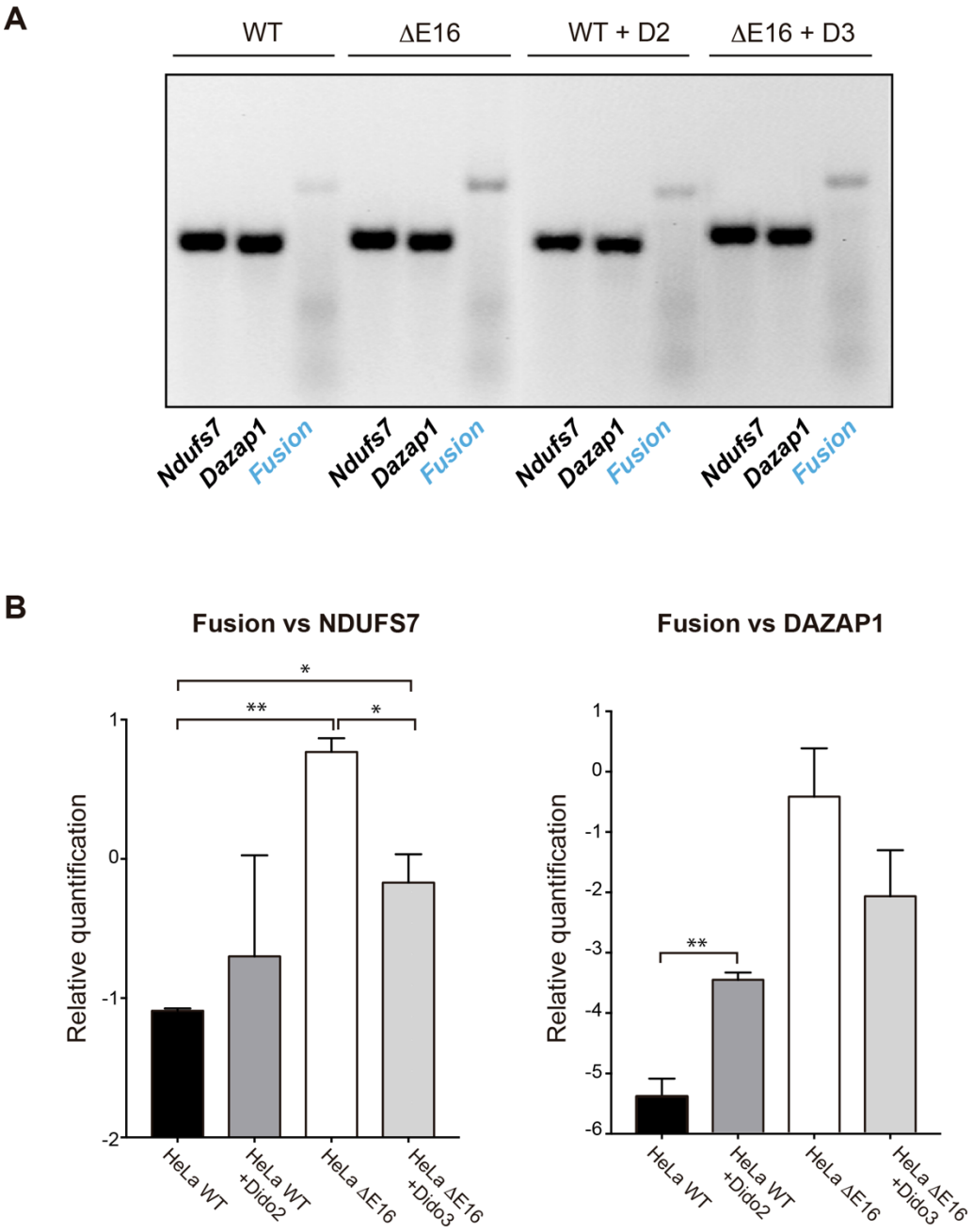



**Figure 32. Complementation with full-length DIDO3 protein in ΔE16 mutants partially restores the normal splicing behaviour**. (A) Reverse transcriptase PCR of *Ndufs7*, *Dazap1* and fusion transcripts in parental and complemented cell lines. (B) RT-qPCR analysis of all samples. The first graph depicts the comparison between the expression levels of the *Ndufs7* gene and the fusion transcripts. The second graph represents the same comparison but using the *Dazap1* gene. In both cases, ΔE16 mutant shows higher fusion levels, while DIDO3 complementation promotes normal *Ndufs7* and *Dazap1* expression. Both graphs depict relative quantification values on Log 2 scale Error bars represent SD from three different samples. *p<0.05; ***p<0.001.

To check SFPQ accumulation in nuclear speckles, the parental and complemented HeLa-FLP cells were treated with the Isoginkgetin (IGK) inhibitor (see **Results Chapter 1.2**). As observed previously, immunofluorescence experiments showed SFPQ accumulation in supramolecular complexes in WT cells treated with this splicing inhibitor (**Figure 33A**). The

proportion of large complexes was reduced in the human ΔE16 cells. Nevertheless, after DIDO3 complementation, speckle amount increased to the WT levels and the phenotype was restored. On the other hand, overexpression of DIDO2 in WT cells diminished the generation of these SFPQ complexes to levels found in the ΔE16 mutants (**Figure 33B**).

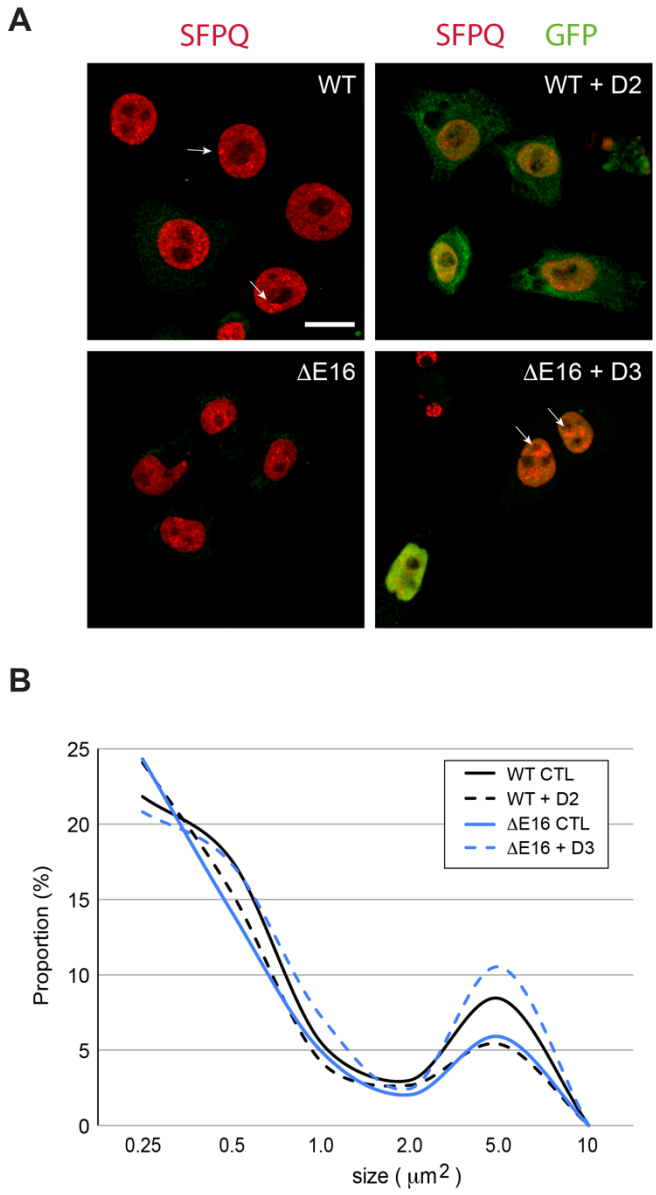



**Figure 33. Complementation of DIDO3 in ΔE16 mutants restored the accumulation of SFPQ in supramolecular complexes**. (A) Immunofluorescence of HeLa-FLP cells labelled with anti-SFPQ (red) and treated with 40 μM IGK shows the formation of supramolecular complexes (arrows). DIDO2 and 3 isoforms are depicted in green due to the eGFP expression. Representative images are shown. Scale bar, 10 μm. (B) Measurements of SFPQ accumulation in supramolecular complexes. The introduction of DIDO3 in HeLa-FLP ΔE16 mutants restored the WT phenotype, while expression of DIDO2 in a WT background prevented complex formation.

In conclusion, the combination of splicing alterations and reduced SFPQ accumulation in supramolecular complexes was confirmed in human ΔE16 cells. In addition, complementation with the full-length DIDO3 isoform mostly restored the WT phenotype, confirming the specificity of the alterations related to the lack of DIDO3.

## Chapter 3. The DIDO3-SFPQ axis regulates alternative splicing

### 3.1. *Dido* mutations reduce SFPQ binding to mRNA

The human model was a very useful tool for complementation approaches and the confirmation of the relationship of DIDO3 to splicing alterations. However, MEFs allow us to work with both types of DIDO3 mutants and to broaden our vision of the DIDO3 role in the splicing process. In particular, the ΔNT and ΔE16 mutations correspond to different protein domains in DIDO3.

Previous results showed that SFPQ is tightly linked to DIDO3. Since *Dido* mutation modulated the effect of splicing inhibitors on SFPQ recruitment and distribution, the same mutations could also compromise SFPQ interaction with RNA. To evaluate this hypothesis, PAR-CLIP experiments as described (Castello, Horos et al. 2013, Takeuchi, Iida et al. 2018) were carried out in MEF cells. We crosslinked SFPQ to the mRNA, isolated the protein-RNA complexes using oligo (dT) beads and quantified the amount of SFPQ recovered from every sample (**Figure 34**). Compared to WT MEF, SFPQ levels were markedly reduced in ΔE16 samples and moderately diminished in ΔNT mutants. Negative controls showed barely any signal (**Figure 34A**). Quantitation of three independent PAR-CLIP experiments revealed that *Dido* exon 16 deletion significantly reduced the SFPQ binding to the RNA molecules (**Figure 34B**).

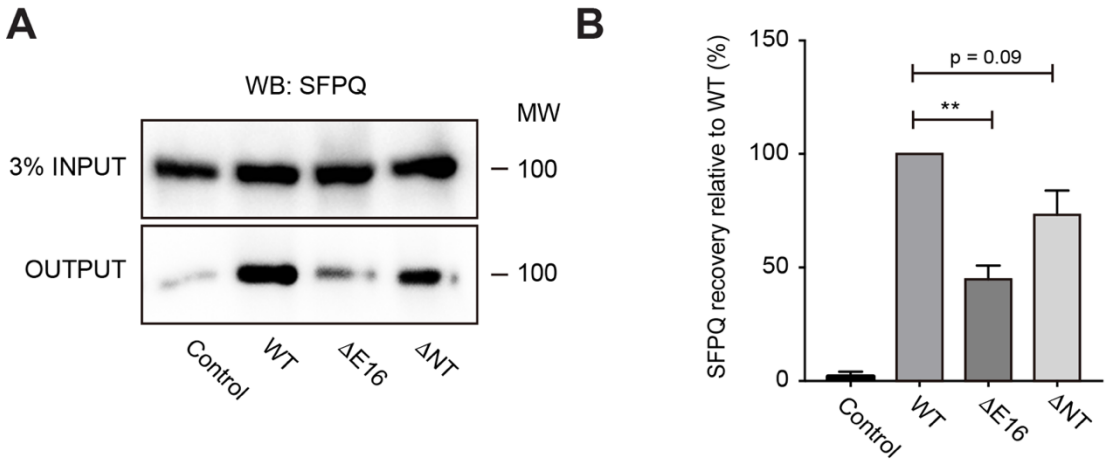



**Figure 34. *Dido* mutations reduce SFPQ binding to mRNA.** (A) Representative image from three independent PAR-CLIP experiments with MEF samples. After crosslinking, polyadenylated RNA molecules and the associated RBPs were isolated. Eluates were loaded into a SDS-PAGE acrylamide gel and separated by molecular weight. Membranes from Inputs and IP samples were developed with an anti-SFPQ antibody. (B) SFPQ recovery was measured by the quantification of band intensity from three PAR-CLIP experiments. The quantifications were performed using Image J software. All numbers were normalized to the WT intensity (100%). Error bars represent SD from three different samples. **p<0.01.

After confirming the effect of *Dido* mutations on SFPQ binding, we tested SFPQ recruitment to specific genes. In this case, Crosslinking and Immunoprecipitation (CLIP) experiments were carried out following a previously described protocol (Ule, Jensen et al. 2003). Here, RNA-protein complexes were photocrosslinked using UV irradiation and SFPQ protein was immunoprecipitated. After protein digestion, the associated RNAs were extracted and quantified using a Bioanalyser (**Figure 35A**). Finally, to identify targets but avoid cDNA library construction, we searched a CLIP-Seq experiment published and available online (Takeuchi, Iida et al. 2018). Once the SFPQ peaks were cross-referenced with our altered genes, we selected one of the most affected genes regarding alternative splicing, *Gas2L1*. Thus, primers for the specific amplification of the exon boundaries corresponding to the SFPQ peak were designed (**Figure 35B**). A reverse transcriptase PCR (RT-PCR) was used to amplify the RNA extracted from inputs and immunoprecipitated samples. A band of correct size for the *Gas2L1* gene was obtained in all inputs. While the WT outputs showed the same band, no amplification was produced by the *Dido* mutants (**Figure 35C**). These results confirmed the defect of SFPQ recruitment to the mRNA when DIDO3 is absent.

Since extensive splicing alterations were found in the ΔE16 mutant and to a lesser degree in the ΔNT mutant, we asked whether the exons affected by these alterations correlated with SFPQ binding deficiency. To further evaluate the SFPQ occupancy on the 5′ end of alternatively spliced exons, the published SFPQ CLIP-Seq data was used (Takeuchi, Iida et al. 2018). We define occupancy as the coverage in CLIP-Seq output corrected for expression level (derived from inputs). First, a list of the significantly altered exons in both *Dido* mutants as compared to WT was generated. Next, the SFPQ occupancy in these exons was compared to a random unaltered population (**Figure 36A**). Exons affected by altered splicing in ΔE16 or ΔNT mutants showed a significantly higher SFPQ occupancy with respect to the random population. In addition, SFPQ occupancy was related to the type of alternative splicing event; we found a positive correlation between exon skipping and the presence of an SFPQ peak. In contrast, exons prone to inclusion showed less SFPQ occupancy (**Figure 36B**).

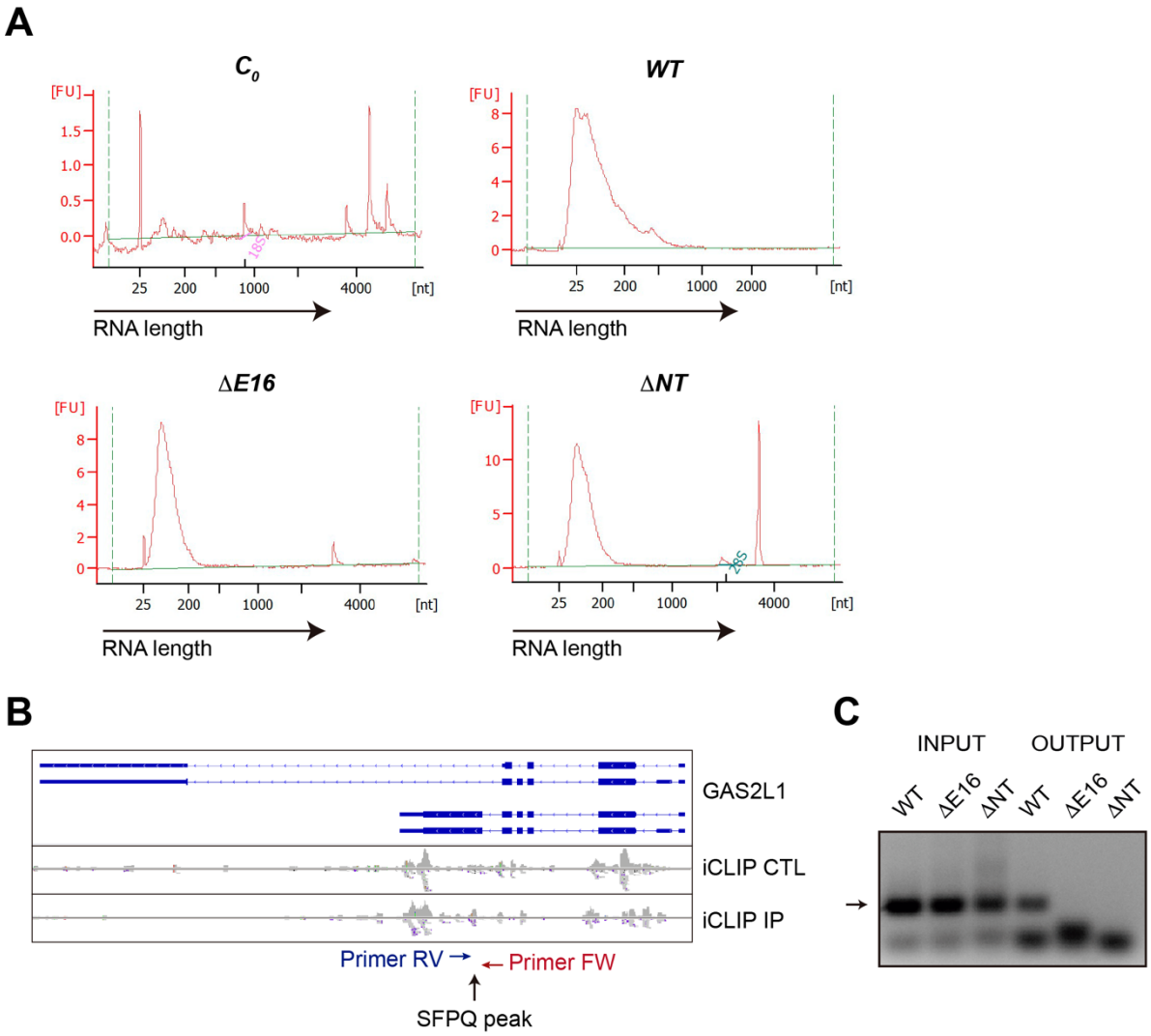



**Figure 35. SFPQ-dependent exons are affected by *Dido* mutations**. (A) Bioanalyser results showing the length and quantity of the RNA extracted from SFPQ CLIP experiment. The vertical axis represents RNA concentration measured in fluorescence units (FU). The horizontal axis represents the length of the recovered RNAs. We obtained detectable amounts of RNA, 50-100 nucleotide in length, from all samples except the negative control ($C_0$). (B) Schematic representation of the mouse *Gas2L1* gene including the iCLIP-Seq analysis (Takeuchi, Iida et al. 2018). Control (CTL) and IP data show specific SFPQ peaks at the *Gas2L1* gene (black arrow). Primers designed for CLIP amplification are shown in colours. (C) While reverse transcriptase PCR produced a band in the WT sample no *Gas2L1* amplification was found in both *Dido* mutants. Representative picture from three different experiments.

Together, these results confirm that exons regulated by SFPQ are the ones more alternatively skipped in ΔE16 and ΔNT mutants. Thus, the lack of DIDO3 reduces SFPQ binding to messenger RNA, affecting the splicing of a group of susceptible exons. Furthermore, exons with similar SFPQ occupancy tend to be more included in the ΔNT mutant but more skipped in ΔE16.

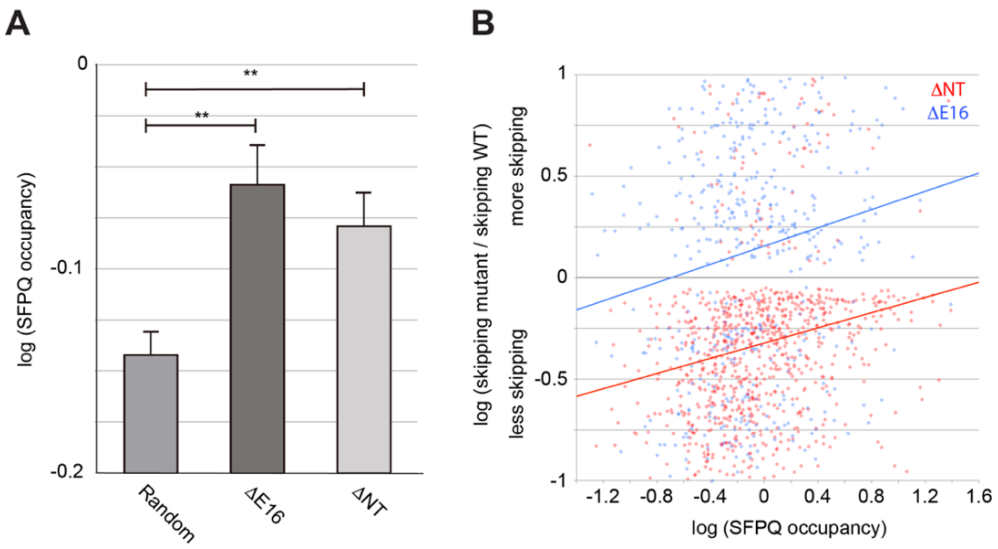



**Figure 36. Exons significantly more skipped in both *Dido* mutants correlate with SFPQ occupancy.** (A) Quantification of SFPQ binding to a random control group or to altered exons in both *Dido* mutants. **p<0.01. Error bars represent standard error of the mean. (B) Positive correlation between SFPQ occupancy and individual exons that are skipped significantly more or less in ΔE16 and ΔNT mutants as compared to WT.

### 3.2. *Dido* mutations alter the preference for upstream 3'SS

According to SFPQ dependence, exons respond differently to *Dido* mutations. Whereas ΔNT increased the exon inclusion events, ΔE16 mostly gave rise to exon skipping (**Figure 20**). Since SFPQ is involved in the proper recognition of the 3'SS at an early stage of the spliceosome assembly (Gozani, Patton et al. 1994), we studied whether the use of these 3'SS indeed differed between the mutants and the WT. In order to test this hypothesis, we first designed a scheme of alternative splicing events in which multiple 3'SS, upstream and downstream, share a common 5' donor site (**Figure 37A**). Based on this scheme, the utilization of each 3'SS relative to the alternative sites was calculated. The name "utilization" thus defines the contribution of one 3'SS to the splicing system, measured on a scale from 0 to 100%. According to the designed scheme, the 3'SS were grouped by distance relative to the common 5'SS, more upstream or downstream than alternative acceptor sites.

First, we found that in a random set of unaltered 3'SS (or exons), 80% of the splicing originated from the upstream 3'SS whereas only 10% used the downstream site (**Figure 37C**). This effect was expected, as it follows the "first come, first served" model (Aebi and Weissmann 1987). However, when we calculated WT utilization for the 3'SS of exons that are affected in *Dido* mutants (altered 3'SS), we saw that their behaviour was different. On

average, upstream sites in affected exons were used less than downstream 3'SS. Thus, exons that will be affected in the *Dido* mutants already show more exon skipping than unaltered exons in WT cells (**Figure 37B and C**).

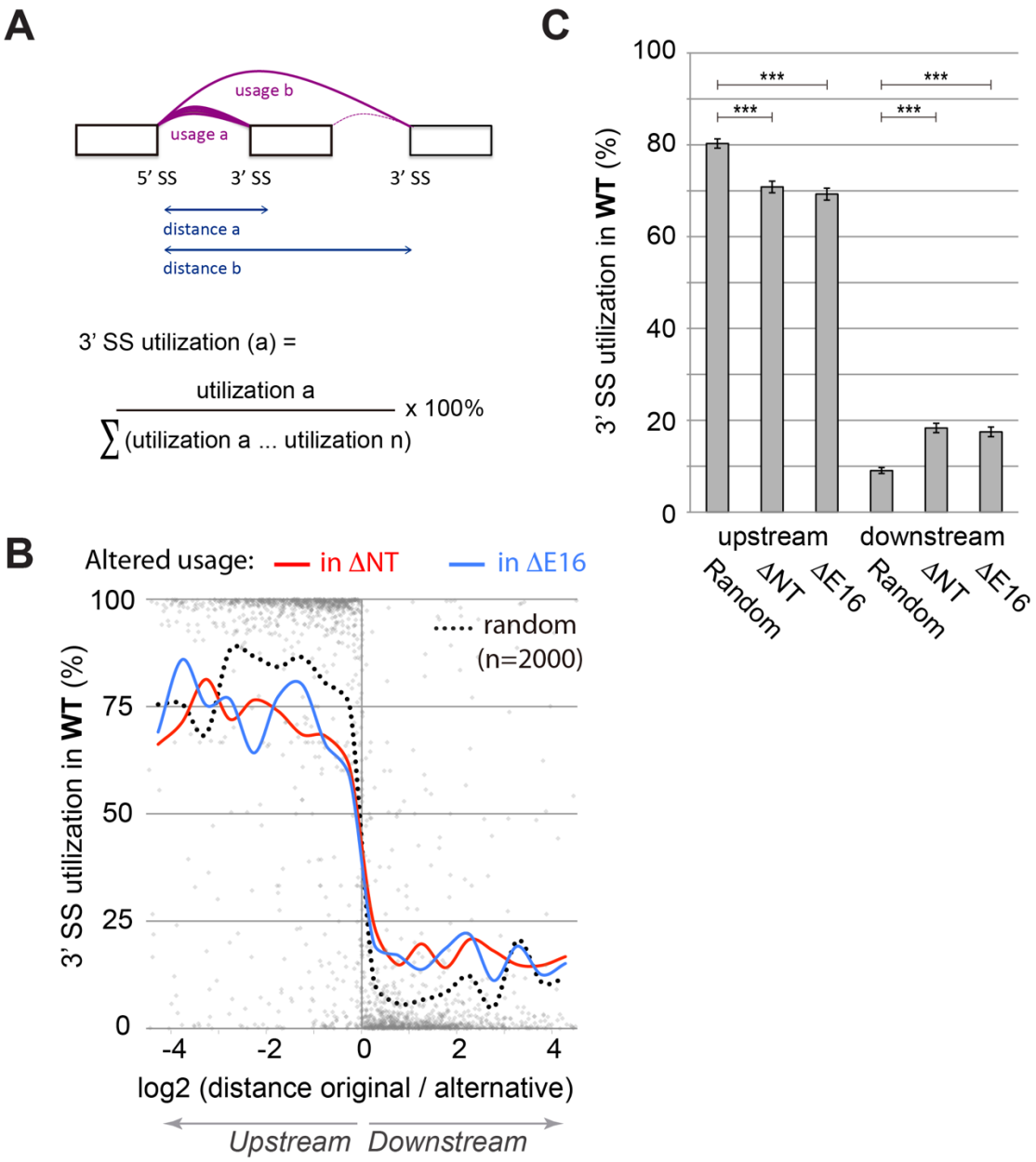



**Figure 37. *Dido* mutants change the normal preference for 3'SS utilization**. (A) Schematic representation of the formula used for utilization calculation. For each 3'SS, the contribution to splicing (utilization) was compared to the alternative sites. (B) The horizontal axis represents the distance of the alternative and the original 3'SS to a common 5'SS. If distance "a" is smaller than distance "b" in panel A, the 3'SS is located upstream and vice versa. The vertical axis represents utilization in WT cells of a random group of 3'SS (dotted line) and utilization in WT cells of 3'SS that are significantly ($p<0.05$) altered in ΔE16 and ΔNT mutants (coloured lines). (C) Comparison of all upstream and downstream 3' SS shows that the differences are significant for the population as a whole. Error bars depict standard deviation. ***$p<0.001$.

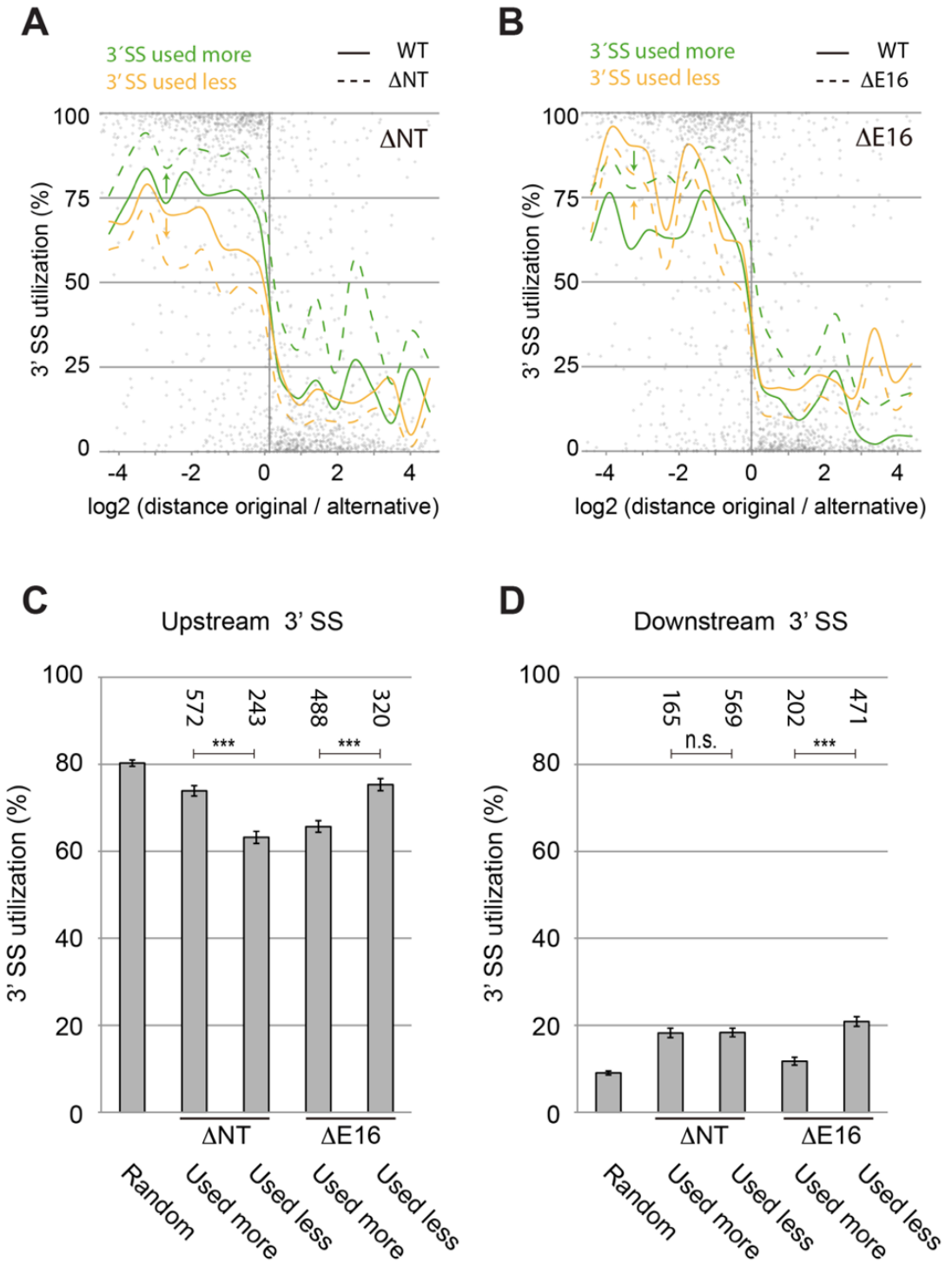



**Figure 38. *Dido* mutations modulate intrinsic preference for upstream 3'SS** (A, B) In these graphs, the 3'SS that were significantly altered in both mutants are separated into two groups. Graphs show WT utilization and mutant utilization for each group. The group that is used more by the mutants is depicted in green and the group that is used less is depicted in orange. Continuous lines represent the WT utilization while the discontinuous lines depict the mutant values. Arrows in both graphs illustrate the effect of 3'SS usage in ΔNT and ΔE16 mutants compared to WT. (C) Differentiation of upstream 3' SS used significantly more or significantly less in each mutant. (D) Differentiation of downstream 3' SS used significantly more or significantly less in each mutant. Error bars depict standard deviation. ***p < 0.001.

Nevertheless, if both mutants are analysed separately, information regarding the real use of upstream and downstream 3'SS can be extracted. So, we divided the list of significantly ($p<0.05$) altered 3'SS in the mutants (depicted in **Figure 37B**) into two groups. One group of 3'SS was used more and the other less as compared to the WT samples. Finally, WT and mutant utilization values were calculated for each 3'SS in each group. These four lists were represented in **Figure 38A** (ΔNT) **and B** (ΔE16).

One of first things that we noticed is that both mutants behaved very differently. Taking a look at the arrows, ΔNT increased utilization of a group of 3'SS that are used frequently in the WT (**Figure 38A**, green lines). Likewise, this mutant decreased utilization of a group of 3'SS that are used infrequently in the WT (**Figure 38A**, orange lines). In contrast, ΔE16 increased utilization of a group of 3'SS that are not often used in the WT (**Figure 38B**, green lines) and supressed 3'SS that are used commonly (**Figure 38B**, orange lines).

Since upstream 3'SS are frequently used in the WT (**Figure 37**), the ΔNT mutant favours exon inclusion. On the other hand, the ΔE16 mutant favours exon skipping, a tendency that we already confirmed in the cases of the *Gas2L1* and *Wdr8* genes (**Figure 23 and 24**). These differences were quantified and the changes in both populations were significant (**Figure 38C and D**). The differences in 3'SS utilization also explained the number of exons that are included or skipped in both mutants (**Figure 21**).

### 3.3. Splice site composition does not influence exon usage

Since splicing alterations in *Dido* mutants are related to the use of individual 3'SS, the next step was the study of the splice site sequence. To test this parameter, we generated a WebLogo representation of differentially used splice sites in ΔE16 and ΔNT mutants (**Figure 39**). Both 3'SS and 5'SS sequences were drawn and compared to a randomized selection of unaltered splice sites. In contrast to what we expected, the composition of the splice sites showed no correlation with 3'SS utilization. In nearly all the cases, no significant differences were recognized between mutants and the random population.

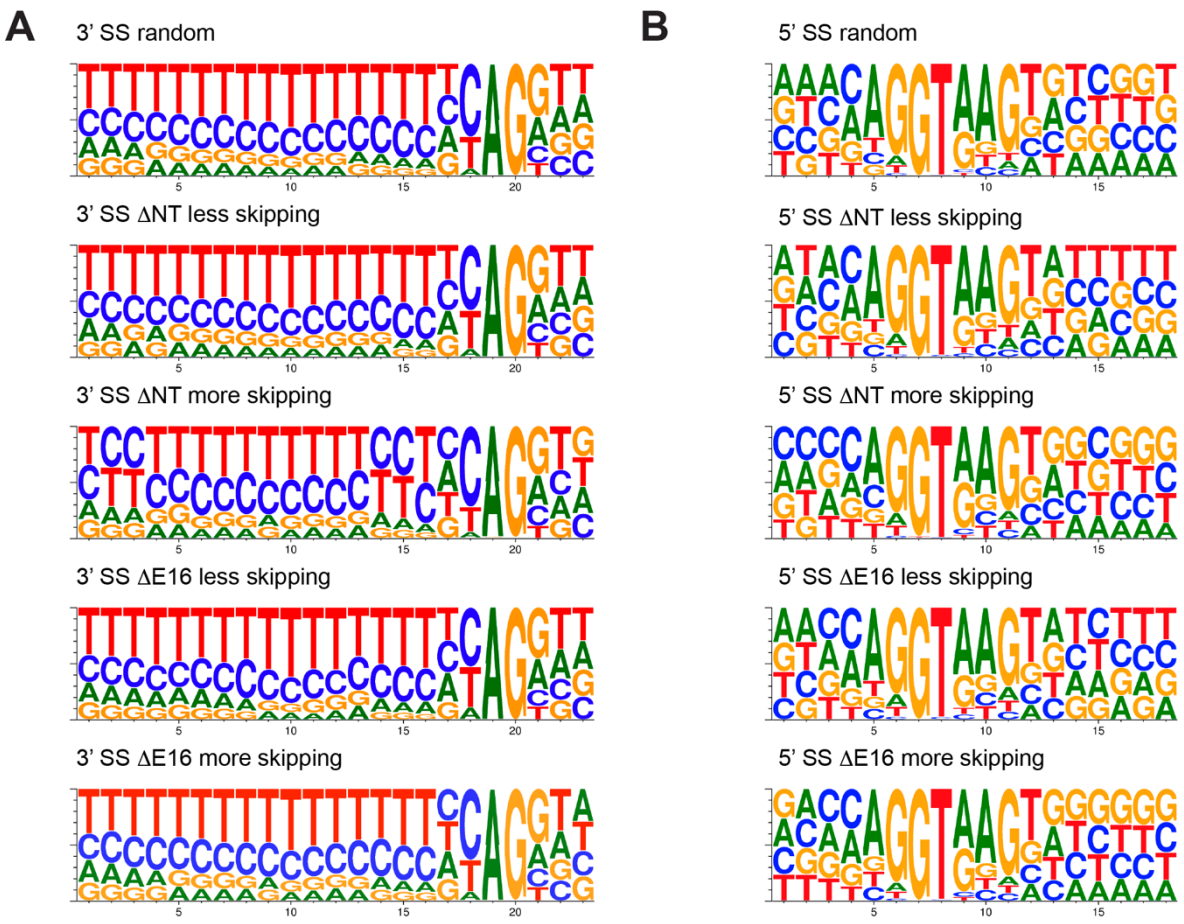



**Figure 39. 5' or 3'SS composition shows no correlation with exon usage in both mutants.** (A, B) Weblogo representations of 3' acceptor site (A) and 5' donor site (B) of differentially used exons in ΔNT and ΔE16 mutants. For 3'SS calculations, we included the PPT sequences. No significant differences were found in any of the cases.

### 3.4. Downstream sequence motifs modulate exon skipping and inclusion

Since splice site composition itself was not implicated in differential exon usage in our mutants, surrounding sequences might influence the splicing alterations. It has been published that regional nucleotide content and short recognition motifs for different splicing factors can modulate splicing in higher vertebrates (Goren, Ram et al. 2006, Amit, Donyo et al. 2012). In order to study the former, we analysed whether the nucleotide composition surrounding alternatively spliced exons in our *Dido* mutants had any bias (**Figure 40**). Using a random selection of unaltered exons as control, we observed that higher Adenine (A) and Thymine (T) content of the surrounding introns correlated with a decrease in exon skipping. The ΔNT mutant showed a larger effect than ΔE16, and barely any differences were found within exons.

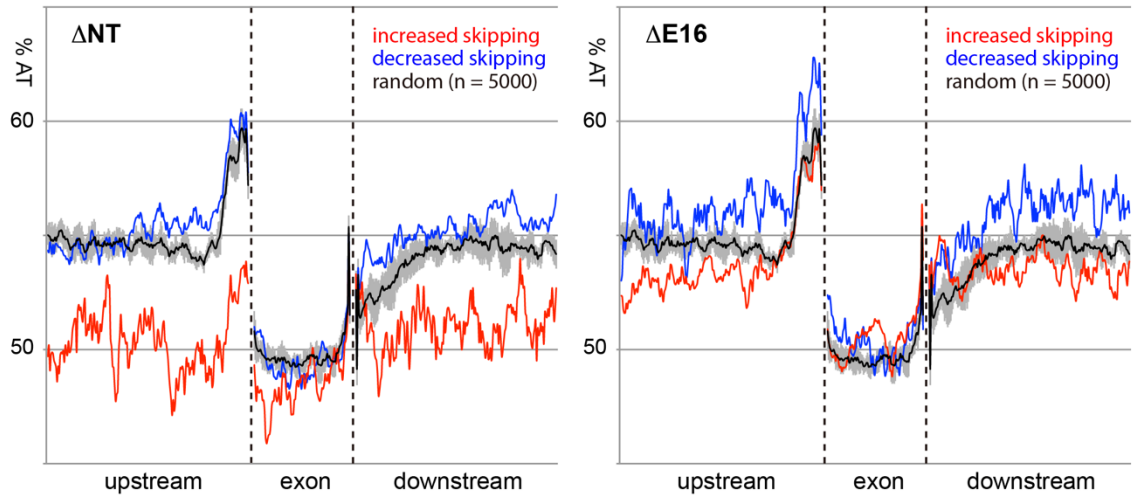



**Figure 40. Nucleotide composition is implicated in alternative splicing.** The AT content (%) was analysed for exons alternatively spliced in *Dido* ΔNT (left) and ΔE16 (right). Black line shows a random selection of unaltered exons that was used as control. Red line represents exons undergoing significantly increased skipping and blue line depicts exons that were less skipped (more included) in *Dido* mutants as compared to WT. 200 bases upstream and downstream of the selected exons were taken into account, and exons were scaled to a 100 base pair window. Error bars depict SD and were derived from a 5 x 1000 random sampling.

After showing that intron composition influences the splicing of the upstream exon, we searched for specific motifs in these regions that could be associated with splicing. Different sequences have been linked to splicing regulation. Specifically, short AT-rich motifs have been associated with pausing of the transcription elongation complexes (Palangat, Hittinger et al. 2004, Saldi, Cortazar et al. 2016). Based on the "window of opportunity" model, if the RNAPII encounters a pause site, this could promote the utilization of upstream 3'SS that are not usually included. We thus searched for all combinations of nucleotide pentamers (5-mers) downstream the altered exon. The frequency of each 5-mer was calculated for both *Dido* mutants, while discriminating between exons undergoing more or less skipping (**Figure 41**). These results indicated that AT-rich sequences, located immediately downstream (<200 bp) of the spliced exon, were predominant in significantly less skipped exons. In addition, the motifs that contained a higher Thymine content were associated with the more included exons in both mutants.

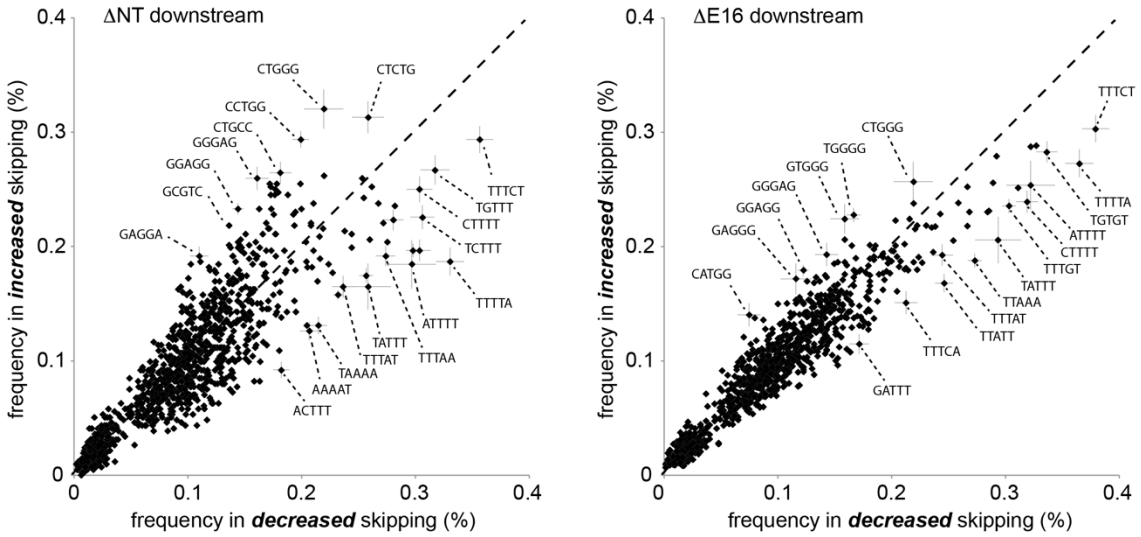



**Figure 41. AT-rich motifs promote exon inclusion in *Dido* mutants.** The scatter plots show the frequency of different 5-mer combinations measured in the first 200 bp downstream of significantly altered exons in *Dido* ΔNT (left) and ΔE16 (right) mutants. A group of 5-mers with significant changes is marked with their associated sequences. The error bars of this group represent standard deviation and were derived from a random selection of 5 x 1000 5-mers. The diagonal discontinuous line indicates no alteration in exon usage.

Although both *Dido* mutants exhibited similar plots, these T-rich motifs contributed more to exon inclusion in ΔNT than in ΔE16. Furthermore, although exon bodies did not presented significant changes in nucleotide composition (**Figure 41**), exonic splicing regulatory sequences (ESRs) have been associated with alternative splicing regulation (Goren, Ram et al. 2006). The ESRs are recognized by different splicing factors such as SRSF1, which has been related with specific purine-rich motifs (Tacke and L.Manley 1995, Cavaloc, Bourgeois et al. 1999, Schaal and Maniatis 1999). Thus, we also searched for 5-mer combinations within alternatively spliced exons (**Figure 42**).

SRSF1 sequences were found in exons that were significantly less skipped in *Dido* mutants. Thus, the presence of these sequences as well as the T-rich motifs correlate with exon inclusion. While T-rich clusters contributed more in the ΔNT mutant, putative SRSF1 motifs had a stronger effect in ΔE16 samples.

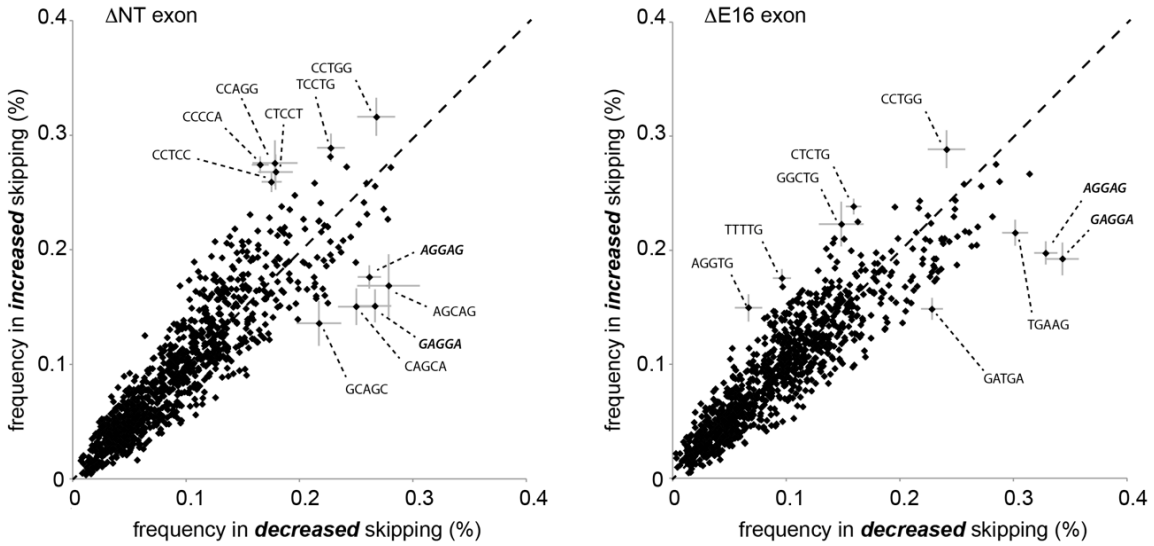



**Figure 42. SRSF1 recognition motifs are associated with exon inclusion in *Dido* mutants.** Both scatter plots show the frequency of 5-mer sequences located in exons that are skipped significantly more (vertical axis) or less (horizontal axis) in ΔNT (left) and ΔE16 (right) mutants. 5-mers undergoing the biggest changes are marked with the associated sequences. SRSF1 recognition motifs are highlighted in bold. The error bars of this group represent standard deviation and were derived from a random selection of 5 x 1000 5-mers. The diagonal discontinuous line indicates no alteration in exon skipping.

### 3.5. T-rich motifs downstream of exons promote their inclusion

In order to confirm that the T-rich motifs were related with the inclusion of some exons in *Dido* mutants, we decided to use the HeLa Flp-In™ T-Rex™ system and generated minigenes with or without such sequences. The model used for minigene generation was the *Ndufs7* and *Dazap1* gene fusion. In our ΔE16 mutant, fusion transcripts were produced from the penultimate exon of *Ndufs7* to the second exon of *Dazap1* (see **Figure 24**). We selected these genes because the behaviour of the splicing junctions differed between mouse and human models even though similar fusion transcripts were found. In the mouse model, the aberrant splicing events started at the end of the last exon of *Ndufs7*, before the 3'UTR, and ended at the second exon of *Dazap1* (**Figure 43A**). However, in the human model, aberrant splicing began at the end of the *Ndufs7* penultimate exon, and finished at the second exon of *Dazap1* (**Figure 43B**). Fusion transcripts were generated in both cases, although only the *Ndufs7* 3'UTR was skipped in the mouse model while the whole last exon and 3'UTR were excised in the human model. Interestingly, when checking the intergenic region between both genes, a long T-rich sequence downstream of the 3'UTR was found only in the mouse model (**Figure 43C**). Since T-rich motifs were related with exon inclusion in our mutants, we

tried to reproduce both mouse and human models by changing only the T-rich intergenic region with a minigene.

Thus, a minigene containing the last two exons of *Ndufs7*, the first and second exons of *Dazap1*, and an intermediate region with the T-rich motif or a random sequence was generated (**Figure 44A**). The complete minigene was cloned into a pcDNA$^{TM}$5-FRT-TO vector and stably transfected into HeLa-FLP WT and ΔE16 cells. Minigene expression was induced with 1 μg/ml of doxycycline for 72 hours and the generation of the fusion transcript was tested by Reverse transcriptase PCR (**Figure 44B**). Although formation of a fusion transcript was observed in all samples, skipping of the whole *Ndufs7* last exon only took place in the ΔE16 mutant and when the T-rich sequence was lacking. The absence of a T-rich motif mimicked the human situation in the *Dido* mutant, while the presence of these sequences promoted inclusion of the *Ndufs7* last exon. Surprisingly, all minigenes produced fusion transcripts even in WT HeLa-FLP cells. This effect might be produced because only two exons from each gene were cloned at a reduced intergenic distance, likely simulating the structure of one single gene. Minigene construction thus favoured the splicing between *Ndufs7* exon 8 and *Dazap1* exon 2 similar to a normal splicing event instead of an intergenic fusion.

Finally, these results confirm that downstream T-rich regions may help to include upstream exons in our *Dido* mutants. This effect could be related with the previously mentioned RNAPII pause sites (Palangat, Hittinger et al. 2004, Saldi, Cortazar et al. 2016). The T-rich motif found near the mouse *Ndufs7* gene is 800 bp downstream of the included exon. However, the T-rich 5-mers identified in both *Dido* mutants were located only 200 bp downstream of the included exon (**Figure 42**). Thus, other factors might influence the use of this T-rich sequence as a pause site in the case of the fusion of *Ndufs7* and *Dazap1*. We hypothesize that the strength of the sequence, since it was a very long motif, and the high content in thymine nucleotides could promote RNAPII arrest.

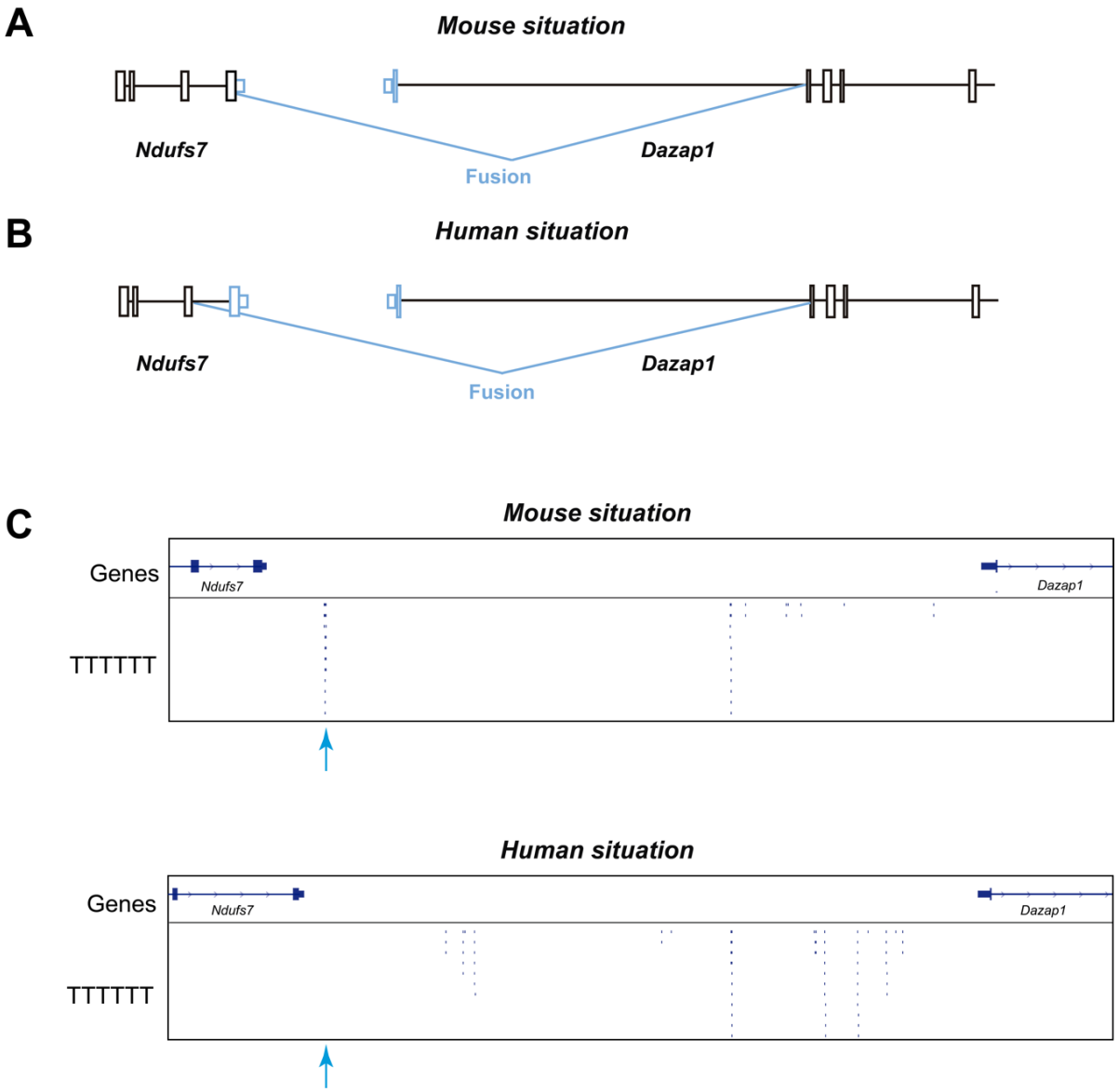



**Figure 43. *Ndufs7* and *Dazap1* fusion transcripts are generated by different splicing events in mouse and human cells.** (A) Schematic representation of the *Ndufs7* and *Dazap1* fusion transcript in mouse cells (MEFs). The splicing event begins at the 3'UTR of *Ndufs7* and ends at the second exon of *Dazap1*. (B) Schematic representation of the *Ndufs7* and *Dazap1* fusion transcript, this time in human (HeLa-FLP) cells. Here, the splicing starts at the *Ndufs7* penultimate exon and again finishes at the second exon of *Dazap1*. (C) IGV visualization of the downstream sequences of *Ndufs7* in mouse and human cells.

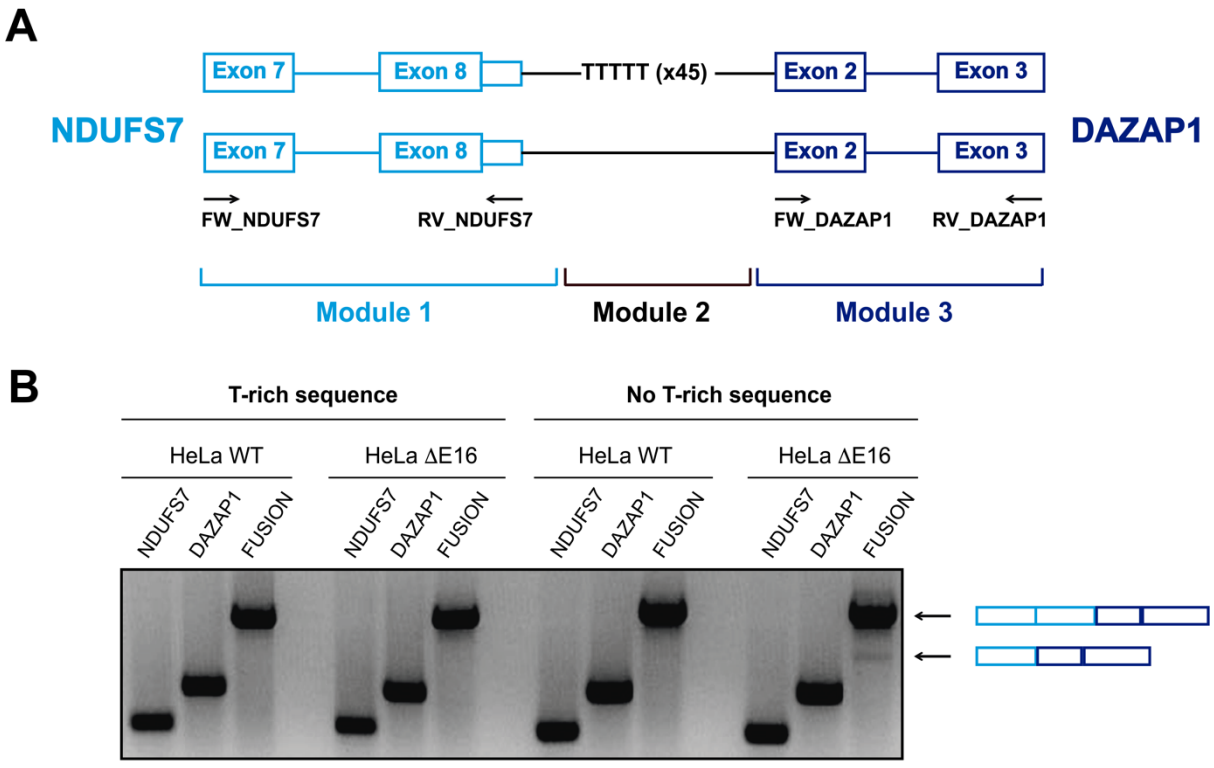



**Figure 44. T-rich sequences promote the inclusion of *Ndufs7* last exon.** (A) Schematic representation of the two cassettes generated for the minigenes, with and without the T-rich region. (B) Reverse Transcriptase PCR experiments show that only in the absence of T-rich sequences downstream of *Ndufs7*, the last exon is skipped. Representative image of three independent experiments.

**Together, the results presented in this thesis project link DIDO3 protein with alternative splicing. The DIDO3 interaction with SFPQ and its role in SFPQ recruitment to messenger RNA connect our protein to U2 snRNP positioning. In addition, the differences found in exon skipping and inclusion seem to depend on 3'SS utilization and SFPQ occupancy. Our data thus indicate the existence of a DIDO3-SFPQ axis acting in alternative splicing regulation. The presence of additional sequences, such as T-rich motifs and SRSF1 recognition sites, affect the splicing pattern in our mutants. This implies an alternative regulatory pathway, likely involving RNA polymerase II.**

# DISCUSSION

## 1. The evolution of DIDO3/exon 16 satisfies a requirement in alternative splicing

RNA splicing is a complex process that is highly conserved among organisms (De Conti, Baralle et al. 2013). Although it was considered for decades as a post-transcriptional process, most recent publications couple splicing and transcription as simultaneous events (Neugebauer 2002, Oesterreich, Herzel et al. 2016, Herzel, Ottoz et al. 2017). Alternative Splicing is a variation of splicing that is prevalent in vertebrates, and allows the generation of multiple isoforms of the same messenger RNA (Chow LT, RE et al. 1977). While the basic principles of splicing are conserved in all eukaryotes, alternative splicing requires additional levels of regulation (Keren, Lev-Maor et al. 2010). One of the most studied mechanisms implicated in this regulation is the correct positioning of the U1 and U2 snRNP on splice donor and acceptor sites. Since U1 interacts directly with the mRNA, 5'SS recognition does not involve helper proteins (Spiluttini, Gu et al. 2010). However, U2 placement requires additional factors. Correct 3'SS recognition by the U2 snRNP is guided by accessory splicing factors, such as U2AF, SR family or SFPQ proteins (Ruskin, Zamore et al. 1988, Patton, Porro et al. 1993, House and Lynch 2008). Thus, the splice acceptor site is a frequent point of interest to study alternative splicing regulation.

In this thesis project, we studied the role of the DIDO3 protein in alternative splicing. We show that DIDO3 acts as a 3'SS modulator through SFPQ recruitment. Although the *Dido* gene encodes for three different isoforms that contain several domains, we focused on DIDO3, the longest protein. This large isoform is only produced in vertebrates, whereas DIDO1 and DIDO2 can be found in all eukaryotes (Rojas, Sanchez-Pulido et al. 2005). No conserved domains have been assigned to the DIDO3-specific region that is encoded by exon 16. However, the smaller isoforms such as DIDO1 and DIDO2 have been described in more detail. The association of these isoforms with histones and the transcription machinery is mediated by the PHD and TFS2M domains, respectively (Gatchalian, Futterer et al. 2013, Futterer, de Celis et al. 2017). Since DIDO3 expands the modular structure of the smaller isoforms, it may simply add another function related to RNA metabolism.

Different lines of evidence connect the DIDO3-specific domain to alternative splicing. Regarding evolution, this type of specialized splicing is extensively used in vertebrates, the same organisms where the DIDO3 full-length protein is found. In lower eukaryotes (insects and yeast), some alternative splicing occurs but is not the dominant form of splicing. BLAST

searches found homologs of SFPQ, one of the major DIDO3 interactors, in chordates and arthropods but not in lower eukaryotes. We thus hypothesize that DIDO3 and SFPQ appeared at the same time and evolved together. The combination of the two may provide an extra level of regulation in the splicing process.

## 2. DIDO3 and SFPQ, two very different but closely related proteins

In this project, we proved that SFPQ interacts specifically with DIDO3, because no binding was found with the shorter isoforms (**Figure 11**). Although this interaction was not assigned to any particular protein domain, the coiled-coil in both DIDO3/exon 16 and SFPQ is a good candidate. A close look into the modular structure of SFPQ finds similarities with the DIDO3 carboxy-terminus. SFPQ belongs to the DBHS or *Drosophila behaviour/human splicing* protein family, together with NONO and PSPC1. This family is implicated in several steps of gene regulation including RNA transcription, splicing, and even RNA transport. The three proteins contain tandem RRMs, a coiled-coil, and a NOPS protein-protein interaction domain (Knott, Bond et al. 2016). SFPQ homodimers interact via their coiled-coil oligomerization motifs. The same domain is also involved in the heterodimerization and oligomerization of the DBHS protein family, facilitating their interaction (Lee, Sadowska et al. 2015). Thus, although the DIDO3 carboxy-terminus is very unstructured and disordered, it contains short regions similar to SFPQ, such as the coiled-coil and an RRM domain. Disordered regions are common in RBPs and splicing factors of recent evolution (Haynes and Iakoucheva 2006, Jarvelin, Noerenberg et al. 2016). In addition, DIDO was found in several proteomics screenings for RNA-interacting proteins, although these analyses did not discriminate what isoform was detected (Baltz, Munschauer et al. 2012, Castello, Fischer et al. 2012). Indeed, our iCLIP experiments suggested that DIDO3 is a bona fide RBP (**Figure 18D**), potentially implicating the RRM domain. These results fit with the idea of a DIDO3 protein that is involved in alternative splicing and uses the protein domain encoded by exon 16 as a link to SFPQ and mRNA. This idea is also supported by Isoginkgetin and PLA experiments, where the lack of this exon 16 reduced the recruitment of SFPQ to spliceosomal complexes (**Figure 13A and 17**). The observation that hnRNP U, another DIDO3 interactor, did not colocalize with DIDO when splicing was inhibited denotes SFPQ specificity (**Figure 13C**).

Several studies have proposed that SFPQ is recruited to the nascent transcript by the RNAPII CTD (Emili, Shales et al. 2002, Rosonina, Ip et al. 2005). However, a later study could not confirm a relationship of SFPQ with the polymerase (Das, Yu et al. 2007); this study proposed that only other splicing factors such as SRSF1 associate with RNAPII CTD. These authors claimed that the SRSF1 protein, together with the U1 snRNP, are recruited via the CTD and positioned close to the mRNA. The CTD seems to compete with an independent pathway that is governed by hnRNP negative factors, to regulate spliceosome assembly (Das, Yu et al. 2007). Whether or not SFPQ is recruited via the CTD, we showed that DIDO3 is one of the sources of SFPQ delivery to RNA transcripts. PAR-CLIP experiments revealed that SFPQ accumulation and binding to mRNA are significantly reduced in our *Dido* mutants, particularly in the one lacking the SFPQ-interacting region (ΔE16) (**Figure 34**). We thus envision that DIDO3 forms a bridge between RNAPII and SFPQ, connecting through the TFS2M domain (Kinkelin, Wozniak et al. 2013, Futterer, de Celis et al. 2017) and carboxy-terminal region, with the objective to place the latter protein close to the nascent RNA (**Figure 45**).

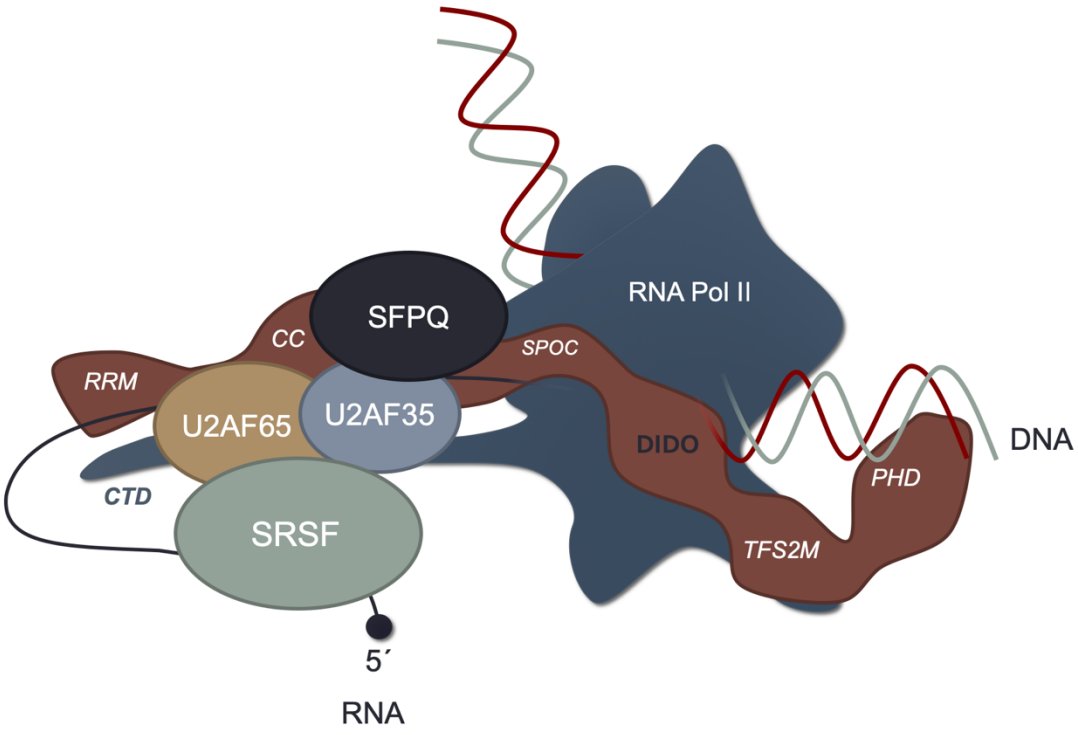



**Figure 45. Schematic model of the DIDO3 full-length protein (brown) associated with RNAPII and SFPQ.** Here, the coiled-coil domain would recruit SFPQ to help the correct positioning of U2 on the 3'SS of nascent RNA. In a parallel pathway, other SF such as SRSF1 assist U2 snRNP and regulate exons independently. The PHD domain binds DIDO3 to the chromatin and the TFS2M and SPOC domains manage RNAPII association. The putative RRM could also be implicated in direct binding of DIDO3 to the RNA.

This model is further supported by PLA experiments that measured the interaction between SFPQ and the RNAPII (**Figure 46**). While SFPQ recruitment to the RNAPII seems to be unaffected in the ΔE16 mutants, the lack of the PHD domain in ΔNT reduces SFPQ interaction with the polymerase. The latter mutant might somehow interfere with an RNAPII-dependent route of SFPQ recruitment.

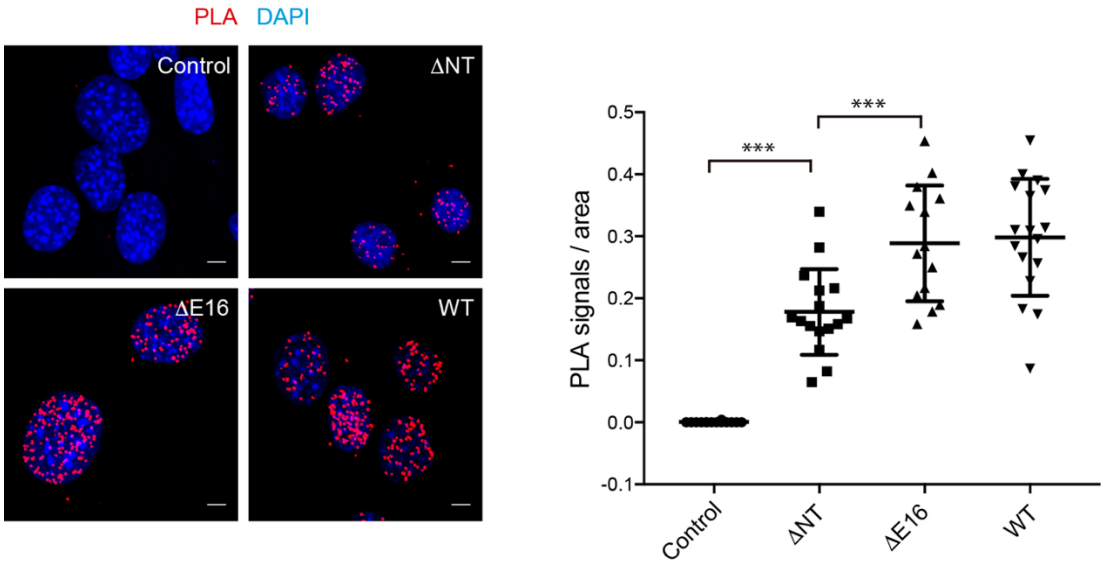



**Figure 46. SFPQ recruitment to RNAPII is reduced in ΔNT mutants.** Immunofluorescence images of Proximity Ligation Assays between SFPQ and RNAPII (left) in WT and *Dido* mutants. Red spots represent proximity between both proteins, while nuclei are depicted in blue. The control sample was labelled with only one antibody. The antibody used for RNAPII detection recognizes the CTD repeat YSPTSPS. PLA signal quantitation (right) revealed a significantly reduced fraction of SFPQ close to RNAPII in ΔNT mutants compared to WT and ΔE16 mutant cells. Error bars represent standard deviation. ***p<0.001.

### 3. *Dido* mutations correlate with SFPQ-associated splicing defects

Since splicing factors are necessary for correct cell development, generation of relevant mutants has always been a difficult process. Indeed, mutations associated with different splicing factors are directly involved in several pathologies and especially in cancer (Oltean and Bates 2014, Singh and Eyras 2017). When factors such as SF3B1, U2AF1 or SRSF2 are mutated, they promote the development of haematological malignancies (Cheng, Luo et al. 2016, Dvinge, Kim et al. 2016). While the constitutive knock-out of NONO and PSPC1 was possible (Li, Li et al. 2014) different attempts to generate SFPQ insufficiency by immunodepletion have been unsuccessful (Patton, Porro et al. 1993, Gozani, Patton et al. 1994). Furthermore, recent studies published the generation of SFPQ null mutants in zebrafish and mouse, but SFPQ silencing in both models was incompatible with long term cell viability (Thomas-Jinu, Gordon et al. 2017, Takeuchi, Iida et al. 2018). Nevertheless, we

were able to delete large portions of *Dido* (ΔNT and ΔE16) while maintaining cell viability (**Figure 14 and 15**). Our mutants allowed the study of splicing defects without disrupting any canonical factor that affected cell survival. Nonetheless, several thousand of splicing alterations were found in both mutants, most of them regarding exon skipping and exon inclusion (**Figure 20**). Since the presence of these alternative splicing events were confirmed in a human model (**Figure 28, 29, 30 and 31**) we assumed a direct relation with the *Dido* mutations.

After having determined that DIDO3 acts by recruiting SFPQ, some kind of correspondence between the lack of SFPQ and DIDO3 was expected. Given the impact of SFPQ deletion, reduced cell viability might compromise RNA sequencing data in these models (Takeuchi, Iida et al. 2018). The same studies however included WT controls that could be analysed. Indeed, evaluation of the SFPQ dependency of the exons with altered splicing in our *Dido* mutants showed a positive correlation (**Figure 36**). Exons skipped significantly more in our mutants correlated with a high SFPQ occupancy, while exons included more presented lower SFPQ association levels. This correlation with SFPQ occupancy fits with the general idea of splicing factor recycling. As mentioned in the introduction, spliceosome formation is a very dynamic process. Many splicing factors need to be re-used for every cycle of splicing, sometimes even in exons from the same gene (Medenbach, Schreiner et al. 2004, Wahl and Lührmann 2015). This recycling process is quite important, because the maintenance of a reservoir of free subunits may decide the inclusion or skipping of certain exons. In line with this hypothesis, suppression of recycling strongly reduces RNA splicing yield, and mostly affects exons containing weak U2-binding sites. A general lack of splicing factor availability impairs proper splice site recognition by the spliceosome. The resulting splicing suppression may even lead to different pathologies and cancer (Han, Park et al. 2017). At least one publication reports that loss of SFPQ mainly affects long genes (Takeuchi, Iida et al. 2018). The sequencing data in this publication show a reduction but not complete absence of SFPQ. Possibly, long genes suffer more because their transcription requirements (a prolonged burst of transcription) easily deplete the remaining SFPQ pool. Although our data show no changes in total SFPQ levels, *Dido* mutations affect recognition of U2-binding sites as evidenced by excessive exon skipping. Thus, instead of global availability, SFPQ recruitment to individual RNA transcripts is diminished.

The group of skipped exons are typical SFPQ targets and can be easily explained by the 3'SS utilization parameter. Exons that already in WT cells showed more skipping than unaltered exons were affected more in the *Dido* mutants (**Figure 37**). Thus, a decrease in SFPQ availability impairs the proper recognition of 3'SS and promotes the skipping of the corresponding exon. However, the analysed data also showed a second group of 3'SS that were used more frequently in our mutants (**Figure 38**). To explain this second group, we again refer to the model where SFPQ is recruited from a shared pool. Thus, when access of SFPQ to one 3'SS subset is reduced in the mutants, its availability for the remaining 3'SS automatically increases. The data agree with this model, because a second group of exons benefit from the increased SFPQ availability and are included more in the mutants than in the WT. SFPQ occupancy in this group was lower than in skipped exons, suggesting low SFPQ dependency or faster processing.

### 4. Downstream motifs are implicated in exon skipping and inclusion

Different publications have described the importance of downstream sequences on splice site selection (Goren, Ram et al. 2006). Thus, parameters different from splice site strength or utilization may explain whether an exon is included or skipped. Exons significantly more included in both *Dido* mutants showed downstream repetitive sequences rich in Adenine (A) and Thymine (T) (**Figure 40**). These AT-rich motifs have been associated with RNA polymerase II pause sites in various studies (Artsimovitch and Landick 2000, Palangat, Hittinger et al. 2004, Nudler 2012). This delay in RNA processing, which sometimes can persist for minutes (Henriques, Gilchrist et al. 2013), affects splice site choice because it extends the so-called window of opportunity (Aebi and Weissmann 1987, Kornblihtt, Schor et al. 2013, Saldi, Cortazar et al. 2016). The general hypothesis indicates that these pause sequences reduce RNAPII speed, thereby facilitating the inclusion of exons with weak or less used 3'SS (Roberts, Gooding et al. 1998, Carrillo Oesterreich, Preibisch et al. 2010).

In our *Dido* mutants, two separate factors are promoting the increased inclusion of some exons: SFPQ availability and AT-rich motifs. As we discussed in the previous section, a reduced SFPQ availability favours the selection of exons that are less SFPQ-dependent. At the same time, AT-rich motifs would pause the RNAPII, thereby promoting inclusion of upstream exons. Together, both situations interfere with the normal splice site selection and

promote the inclusion of a subset of exons included less frequently in the WT. This effect was also found in other organisms with *Dido* homologs, such as *D.melanogaster.* The PPS protein is a DIDO homolog that favours skipping of the *Sxl* exon that determines male sex. However, in the absence of PPS, inclusion of this exon could be explained by an RNAPII pause in the AT-rich stretches found downstream the exon.

We generated minigenes containing these pausing sequences in order to confirm the effect in exon inclusion (**Figure 44A**). The *Ndufs7* and *Dazap1* genes showed different fusion transcripts in human and mouse ΔE16 mutants. Since the most striking difference was the presence of a long T-rich motif downstream the 3'UTR of *Ndufs7* (**Figure 43**), they were good candidates for minigene design. As expected, the presence of the alleged pause site always promoted the inclusion of the *Ndufs7* last exon and thereby confirmed the effect of the T-rich sequence (**Figure 44B**).

In addition, a 5' cryptic splice site is located at the end of the *Ndufs7* last exon, exactly at the place where the splicing event of the fusion transcript starts (**Figure 47**). Both human and mouse genes contain this U1 snRNP recognition site, but only the mouse model includes the exon. We hypothesize that the pause imposed by the T-rich region creates a window of opportunity for U1 to recognise the cryptic 5'SS. Due to a mechanism termed telescripting (Kaida, Berg et al. 2010, Berg, Singh et al. 2012), U1 binding favours the U2 positioning at the upstream 3'SS and inhibits PAS recognition by the polyadenylation machinery. The combination of the two actions promotes exon definition of the *Ndufs7* terminal exon and subsequent splicing to the next available 3' acceptor site. This latter splice site is the one located in the second exon of *Dazap1*. Although the cryptic 5'SS is encoded by the human genome, absence of the pause site precludes its recognition.

Additionally, our *Dido* mutants show a relationship with splicing factors other than SFPQ. In the mutants, many exons that were included significantly more contained SRSF1 recognition sites (**Figure 40**). Since the SRSF1-RNAPII pathway has been previously described (Tian and Kole 2001, Pandit, Zhou et al. 2013), we hypothesize that exons using this alternative pathway could benefit from the changes in SFPQ availability in the ΔE16 mutant. In the end, different pathways that control 3'SS usage converge on the U2 snRNP.

The observation of SRSF1 recognition motifs in our sequencing data thus indicates that this pathway competes with SFPQ for U2 snRNP positioning.

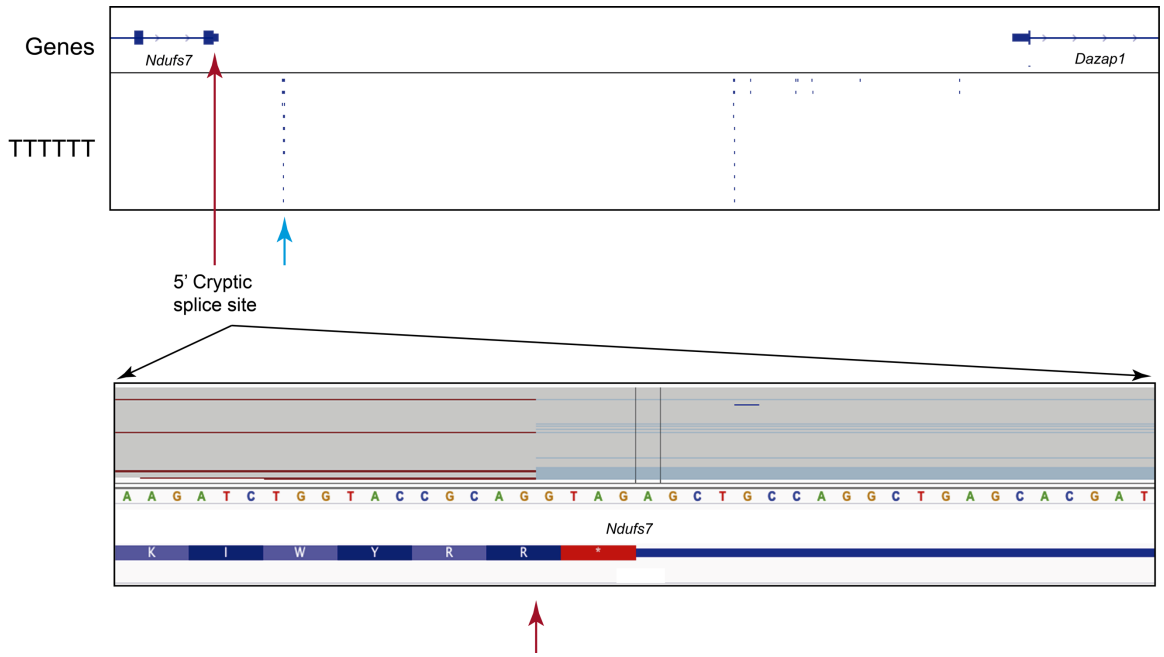



**Figure 47. A cryptic 5'SS is located at the end of the *Ndufs7* last exon.** Schematic representation of mouse *Ndusf7* and *Dazap1* genes depicts the 5' cryptic splice site and the T-rich region. In the box below, the region of the *Ndufs7* last exon and 3'UTR was magnified. The light blue lines and the red arrow depict the cryptic 5'SS.

## 5. Lack of histone binding reduces the DIDO3 reservoir at the TSS

Although the lack of exon 16 reduces SFPQ accessibility for the splicing process, the ΔNT mutant still contains this region. We assumed that the ability to recruit SFPQ is intact in this mutant, according to the mild defects observed in splicing and related PAR-CLIP experiments. In the ΔNT mutant, exons 3 and 4 were eliminated, affecting the binding to the H3K4Me3 (Futterer, Campanero et al. 2005, Gatchalian, Futterer et al. 2013). Thus, the connexion with splicing could be conserved but the association with the chromatin is disrupted. Cases similar to the ΔNT mutant have been reported in the literature. For example CHD1, a modular protein with domains that perform a function comparable to DIDO3. This protein, too, is implicated in transcript elongation and termination in yeast, drosophila and mammals (Sims, Belotserkovskaya et al. 2004). CHD1 associates with the U2 snRNP component SF3A and contains a chromodomain that promotes H3K4Me3 binding. Like DIDO3, this protein accumulates at the Transcription Start Site (TSS) via histone binding and recruits U2 snRNP to the spliceosome, enhancing the efficiency of splicing and transcription

elongation (Sims, Millhouse et al. 2007). In the case of the ΔNT mutant, the lack of a PHD domain impairs histone binding and may prevent accumulation of DIDO3 at the TSS of the genes. Indeed, the reduction of DIDO3 availability affects downstream SFPQ recruitment and the amount of this protein that binds to the mRNA (**Figure 34**).

Since the H3K4Me3 mark is associated with active transcription (Ruthenburg, Allis et al. 2007), the associated chromatin is subjected to continuous passes by RNAPII. During mitotic division, the chromatin is highly condensed and transcription is suppressed (A. and I. 2005). When mitosis is completed, chromatin returns to a more relaxed state and transcriptional programs, including splicing, are rapidly re-initiated. Because H3K4Me3 is closely associated with RNAPII transcription, proteins that recognize this mark are also expelled from the chromatin in mitosis (Gatchalian, Mora Gallardo et al. 2016). Apparently, the bulk of proteins bound to the TSS can interfere with chromatin condensation. Thus, phosphorylation of serine and threonine residues next to prominent methylations ensure the exclusion of histone readers and liberate the chromatin for condensation (van Wely, Mora Gallardo et al. 2017). At the end of mitosis, this situation is reverted. Transcription resumes as soon as possible in the daughter cells, in order to preserve cell identity. To restore gene expression after mitosis, the H3K4Me3 mark, through histone-readers, acts as an active recruiter of transcriptional complexes. While the H3K4Me3 mark is maintained in the chromatin during mitosis, reader proteins are completely excluded. The exclusion of histone readers could expose the H3K4Me3 mark to demethylases, and thus change gene expression patterns. The rapid and massive phosphorylation on adjacent residues at the onset of mitosis thus prevents binding of transcription-related proteins and demethylases alike. As soon as chromatin relaxes and other histone marks such as phosphorylation are removed, histone readers rapidly bind to H3K4Me3. This accumulation at the TSS of active genes initiates the correct gene expression programs (Gatchalian, Mora Gallardo et al. 2016). Some of these histone-readers are proteins with PHD domains, such as ING1, TAF3, DIDO3 or PHF8. The implication of DIDO3 as an histone reader, places our protein already close to the DNA. Given the interplay between H3K4Me3 and transcription, recognition of this mark by DIDO already implicates the protein in this process. Importantly, the H3K4Me3 mark is compressed close to the TSS in non-spliced genes but extends far into the first intron in spliced genes (Bieberstein, Carrillo Oesterreich et al. 2012). We thus assume that a rapid placement of DIDO3 at the 5′ end of genes could help in the generation of a DIDO3 reservoir for subsequent

RNAPII passes. This situation is impaired in the ΔNT mutant since the interaction with the $H3K4Me3$ is disrupted, resulting in the observed splicing defects.

### 6. Splicing alterations might provoke the phenotypes associated with *Dido* mutations

Several phenotypes have been associated with the lack of distinct DIDO regions. Since the deletion of the full-length protein is incompatible with life, different mutants were generated over the years. Absence of the N-terminal part, including the PHD domain, produced mitotic spindle defects (Trachana, van Wely et al. 2007), cell division problems and chromosome instability (Guerrero, Gamero et al. 2010, Martinez and van Wely 2011). In addition, haematological myeloid neoplasms have been reported in mice carrying the ΔNT mutation (Futterer, Campanero et al. 2005). On the other hand, constitutive absence of exon 16 is lethal at embryonic stage 8.5 of gestation (Futterer, Raya et al. 2012). In this case, the lack of exon 16 affected embryonic stem cell (ESC) differentiation (Futterer, de Celis et al. 2017) and impaired iPSC reprogramming (unpublished work). Because of the severity of constitutive deletion, conditional exon 16 knock-outs were generated (**Figure 15**). In *Dido* mutants that survive, a wide variety of alterations is observed. Different types of ciliopathies for example, brain malformations and neuromuscular and behavioural alterations have been associated with a ΔNT/ΔE16 mutant (Villares, Gutierrez et al. 2015). More recently, lack of exon 16 in the hematopoietic lineage was shown to compromise HSCs differentiation, affecting the B cell lineage (unpublished work).

Since $DIDO3$ is a modular protein, the diversity of these phenotypes could be explained due to the wide range of interactions found for DIDO. However, alterations in splicing may be another explanation for the different processes that seem to be regulated or affected by *Dido* mutations. Although the final phenotypes affect apparently unrelated cell types, the large number of genes that show altered splicing could partially overlap with multiple cell-specific expression patterns. In addition, the generation of aberrant mRNA transcripts as a consequence of splicing alterations may lead to the activation of interferon-related genes. Usually, these responses are similar to a viral infection (Zevini, Olagnier et al. 2017). Indeed, gene ontology analyses showed several pathways upregulated in our mutants,

including responses to interferon and defensive strategies against virus or viral production (**Figure 26**).

Mis-regulation of genes that are affected in our mutants may lead to different outcomes. It is important to notice that the splicing alterations that we found, were not "black or white". In most of the cases, a normal transcript was also produced, and only a percentage of the splicing products resulted in aberrant messenger RNAs. However, in some cases, this small percentage of altered transcripts could be sufficient to affect gene function. Here, the case of *Ndufs7* and *Dazap1* was a clear example of such changes in functionality. In ΔE16 mutants, a fusion transcript between both genes was generated. The splicing event between consecutive genes skipped the last exon of *Nudfs7* and the first exon of *Dazap1* in human cells (**Figure 29**). Since the fusion transcript was in frame, a fusion protein was also produced (**Figure 30A**). Interestingly, NDUFS7 is a cytoplasmic protein whereas DAZAP1 remains inside the nucleus. Due to the fusion protein, human ΔE16 cells showed a higher concentration of NDUFS7 in the nucleus compared to WT cells (**Figure 30B**). Here, this aberrant splicing event was affecting the protein localization and possibly its function. Indeed, in the case of DAZAP1, the two RRM domains located at the N-terminal region are disrupted. Since DAZAP1 is an RBP implicated in mammalian development and spermatogenesis, the lack of RNA interaction could affect its function on splicing itself.

Together, our data indicate a role for DIDO3 in the regulation of alternative splicing. A DIDO3-SFPQ axis controls the 3'SS selection in a subset of exons that are highly SFPQ-dependent. In parallel, the RNAPII-SRSF1 axis could promote the inclusion of another group of exons by the help of exon recognition motifs. Both axes could benefit from downstream pause sites. Since DIDO3 binds to SFPQ and RNAPII at the same time, both pathways cooperate to promote effective 3'SS utilization and increase the efficiency of the splicing process. In addition, DIDO3 accumulates at the start site of active genes through the PHD domain, generating a reservoir to be used after transcription initiation. Since splicing is mostly produced co-transcriptionally, DIDO3 travels along the DNA bound to RNAPII, in the meantime recruiting SFPQ to the mRNA. When exon 16 is eliminated, the connexion with SFPQ disappears and therefore the regulation by the DIDO3-SFPQ axis. Thus, exon skipping is increased due to altered 3'SS utilization (**Figure 48**). On the other hand, when the PHD domain is removed, the DIDO3 reservoir is absent, reducing the amount of DIDO3 ready to

recruit SFPQ. This situation also promotes exon skipping of SFPQ-dependent exons but can be overcome by RNAPII pausing (**Figure 49**).

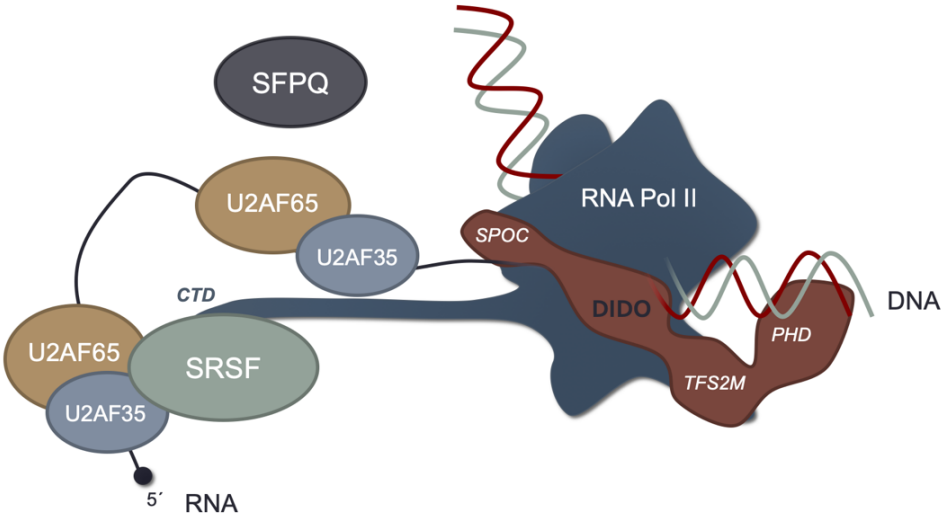



**Figure 48. Schematic model of the Δ16 mutant in RNA metabolism.** The lack of interaction with SFPQ impairs correct delivery of this splicing factor to the 3'SS of a subset of exons. Although U2 snRNP is still recognizing the 3'SS, alternative helper proteins such as SRSF1 may contribute.

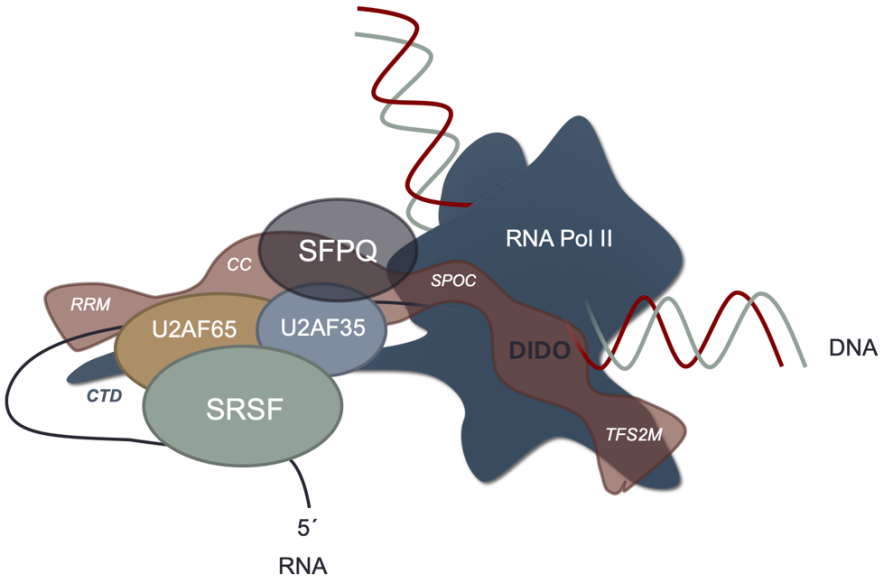



**Figure 49. Schematic model of the ΔNT mutant in RNA metabolism.** The absence of the PHD domain eliminates the connexion with histones and consequently with the DNA. DIDO3 cannot be accumulated at the TSS of the genes and less SFPQ is delivered to the nascent RNA. Consequently, pathways that support posterior recruitment such as RNAPII pause sites could be implicated in the correct splicing of exons.

An important consequence of these models is the inferred competition between DIDO2 and DIDO3. Since these isoforms share the TFS2M domain but not the SFPQ interaction, binding of DIDO2 to RNAPII could limit the possibility to recruit SFPQ. RNA sequencing of heterozygous WT/ΔE16 MEFs indeed revealed an intermediate level of exon skipping (unpublished data). The exact mechanisms that regulate the production of DIDO isoforms in mammals remain to be described. The termination of each isoform in a bleeding exon however implicates telescripting in DIDO isoform selection. A surprising albeit unproven consequence is that DIDO isoforms could functionally link U1 and U2 snRNPs through this mechanism.

# CONCLUSIONS

1. The *Dido* exon 16 that corresponds to the DIDO3 isoform is necessary for SFPQ interaction. The importance of a coiled-coil domain for SFPQ function and presence of a similar domain in DIDO3/exon 16 indicates that the interaction could be mediated by this region.

2. DIDO3 recruits SFPQ to the messenger RNA and *Dido* mutations alter subnuclear SFPQ distribution when splicing is impaired. Given the effect of the Isoginkgetin inhibitor on splicing, DIDO3 and SFPQ seem to act upstream of spliceosome assembly.

3. Elimination of the *Dido* exon 16 (ΔE16) causes thousands of alternative splicing defects, which have been verified in mouse and human models. Most of these events involve alterations in exon skipping. As the absence of the exon 16 reduces SFPQ recruitment, this might impair positioning of the U2 snRNP on the 3'SS.

4. Deletion of the exons 3 and 4 of *Dido* (ΔNT) also produces thousands of alternative splicing defects in a mouse model. However, the number of alterations in exon skipping is lower and the amount of exon inclusion events increases compared to the ΔE16 mutant. Since the exon 16 is still intact, the availability of SFPQ is higher than in the ΔE16 mutant. However, absence of the PHD domain might reduce DIDO3 accumulation at the start site of the genes. Reduced DIDO3 presence on active genes, too, affects SFPQ recruitment, but its impact is lower as compared to exon 16 deletion.

5. Interferon-related pathways are upregulated in both *Dido* mutants. The activation of interferon-related genes and viral defense responses is thus associated with the production of aberrant RNA transcripts.

6. Splicing alterations in human *Dido* mutants are partially corrected by complementation with the full-length DIDO3 protein, to a situation that resembles the WT phenotype. Since DIDO2 overexpression produces alterations similar to DIDO3 absence, the shared PHD and RNAPII-linked domains could produce competition between both isoforms.

7. SFPQ-dependent exons are affected most by the lack of the DIDO3-SFPQ interaction. In these exons, the downstream 3'SS are more utilized than the upstream sites. This change in splice site utilization promotes an increase in exon skipping.

8. Downstream sequences promote exon inclusion under conditions of impaired SFPQ recruitment. T-rich regions, which are associated with RNAPII pause sites, preferentially

facilitate the inclusion of exons that are less SFPQ-dependent in the ΔNT mutant. Exonic SRSF1 recognition motifs promote the inclusion of exons in the ΔE16 mutant, by facilitating their 3'SS utilization.

**9.** A likely scenario for DIDO3 indicates that histone binding accounts for accumulation of the protein at the start site of genes. During transcription, dynamic association with RNAPII positions DIDO3 near the nascent RNA, thereby promoting SFPQ delivery to the messenger RNA and favouring proper recognition of the 3'SS.

# CONCLUSIONES

1. El exón 16 de *Dido*, que corresponde a la isoforma DIDO3, es necesario para la interacción con SFPQ. La importancia de un dominio *coiled-coil* para la función de SFPQ y la presencia de otro dominio similar en el exón 16 de la proteína DIDO3 indica que la interacción entre ambas proteínas podría estar mediada por esta región.

2. DIDO3 recluta a la proteína SFPQ al ARN mensajero y las mutaciones de *Dido* modifican la distribución subnuclear de SFPQ cuando el proceso de *splicing* está comprometido. Teniendo en cuenta el efecto del inhibidor Isoginkgetin en el *splicing*, DIDO3 y SFPQ parecen actuar en un paso previo al ensamblamiento del *spliceosoma* .

3. La eliminación del exón 16 de *Dido* (ΔE16) genera miles de defectos a nivel de *splicing* alternativo, que han sido validados tanto en un modelo de ratón como humano. La mayoría de estos eventos suponen alteraciones en el *skipping* de exones. Dado que la ausencia del exón 16 reduce la disponibilidad de SFPQ, esto podría comprometer el correcto posicionamiento del U2 snRNP a los sitios 3′ de aceptación.

4. La deleción de los exones 3 y 4 de *Dido* (ΔNT) también genera miles de defectos a nivel de *splicing* alternativo en un modelo de ratón. Sin embargo, comparado con el mutante ΔE16, se han observado un menor número de alteraciones relacionadas con el *skipping* de exones mientras que la inclusión aumentaba. Dado que en este mutante el exón 16 permanece en la proteína, la disponibilidad de SFPQ es mayor que en mutante ΔE16. Aún así, la carencia del dominio PHD podría reducir la acumulación de DIDO3 en el sitio de inicio de los genes, lo que afectaría a su vez al reclutamiento de SFPQ, pero en menor medida que en el mutante carente del exón 16.

5. En los mutantes de *Dido*, las vías relacionadas con la respuesta a interferón se encuentran sobre activadas. La activación de genes de esta vía y de respuesta a la defensa viral está asociada a la producción de transcritos de ARN aberrantes.

6. Las alteraciones de *splicing* en los mutantes humanos de *Dido* son parcialmente corregidas por la complementación con la proteína completa de DIDO3, observándose una situación más parecida al fenotipo silvestre. Dado que la sobreexpresión de la isoforma DIDO2 produce alteraciones similares a la ausencia de DIDO3, los dominios compartidos PHD y TFS2M podrían estar generando una situación de competición entre ambas isoformas.

7. Los exones dependientes de SFPQ son los más afectados por la ausencia de interacción entre DIDO3 y SFPQ. En estos exones, los sitios 3' de *splicing* localizados *downstream* se usan más que los sitios *upstream*. Este cambio en el uso de los sitios 3' está asociado a un aumento en el *skipping* de exones.

8. Las secuencias *downstream* promueven la inclusión de exones en condiciones en las que el reclutamiento de SFPQ está afectado. Regiones ricas en Timina, asociadas a sitios de pausa de la ARN Polimerasa II, facilitan la inclusión de los exones que son menos dependientes de SFPQ, especialmente en el mutante ΔNT. Motivos de reconocimiento de SRSF1 localizados en exones promueven la inclusión de exones en el mutante ΔE16, facilitando el uso de sus sitios 3' de *splicing*.

9. El escenario más probable para la proteína DIDO3 indica que la unión con la histona produce la acumulación de la proteína en los sitios de inicio de los genes. Durante la transcripción, la asociación dinámica con la ARN Polimerasa II posiciona a DIDO3 cerca de la cadena naciente de ARN, promoviendo el reclutamiento y entrega de SFPQ al ARN mensajero y favoreciendo el correcto reconocimiento de los sitios 3' de *splicing*.

# REFERENCES

A., K. and T. I. (2005). "Histone modifications defining active genes persist after transcriptional and mitotic inactivation." The EMBO Journal **24**(2): 347-357.

Aebi, M. and C. Weissmann (1987). "Precision and orderliness in splicing." Trends in Genetics **3**: 102-107.

Ahn, S. H., M. Kim and S. Buratowski (2004). "Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing." Mol Cell **13**(1): 67-76.

Amit, M., M. Donyo, D. Hollander, A. Goren, E. Kim, S. Gelfman, G. Lev-Maor, D. Burstein, S. Schwartz, B. Postolsky, T. Pupko and G. Ast (2012). "Differential GC content between exons and introns establishes distinct strategies of splice-site recognition." Cell Rep **1**(5): 543-556.

Artsimovitch, I. and R. Landick (2000). "Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals." PNAS **97**: 13.

Ashfield, R., P. Enriquez-Harris and N. J. Proudfoot (1991). "Transcriptional termination between the closely linked human complement genes C2 and factor B: common termination factor for C2 and c-myc?" EMBO J **10**(13): 4197-4207.

Au, F. K., Y. Jia, K. Jiang, I. Grigoriev, B. K. Hau, Y. Shen, S. Du, A. Akhmanova and R. Z. Qi (2017). "GAS2L1 Is a Centriole-Associated Protein Required for Centrosome Dynamics and Disjunction." Dev Cell **40**(1): 81-94.

Auboeuf, D., M. Carmo-Fonseca, J. Valcarcel and G. Biamonti (2012). "Alternative splicing and cancer." J Nucleic Acids **2012**: 363809.

Baltz, A. G., M. Munschauer, B. Schwanhausser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich and M. Landthaler (2012). "The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts." Mol Cell **46**(5): 674-690.

Barash, Y., J. A. Calarco, W. Gao, Q. Pan, X. Wang, O. Shai, B. J. Blencowe and B. J. Frey (2010). "Deciphering the splicing code." Nature **465**(7294): 53-59.

Barbosa-Morais., N. L., M. Irimia., Q. Pan., H. Y. Xiong., S. Gueroussov., L. J. Lee., V. Slobodeniuc., C. Kutter., S. Watt., R. Çolak., T. Kim., C. M. Misquitta-Ali., M. D. Wilson., P. M. Kim., D. T. Odom., B. J. Frey. and B. J. Blencowe (2012). "The Evolutionary Landscape of Alternative Splicing in Vertebrate Species." Science.

Batsche, E., M. Yaniv and C. Muchardt (2006). "The human SWI/SNF subunit Brm is a regulator of alternative splicing." Nat Struct Mol Biol **13**(1): 22-29.

Behzadnia, N., M. M. Golas, K. Hartmuth, B. Sander, B. Kastner, J. Deckert, P. Dube, C. L. Will, H. Urlaub, H. Stark and R. Luhrmann (2007). "Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes." EMBO J **26**(6): 1737-1748.

Behzadnia, N., K. Hartmuth, C. L. Will and R. Luhrmann (2006). "Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP." RNA **12**(9): 1738-1746.

Berg, M. G., L. N. Singh, I. Younis, Q. Liu, A. M. Pinto, D. Kaida, Z. Zhang, S. Cho, S. Sherrill-Mix, L. Wan and G. Dreyfuss (2012). "U1 snRNP determines mRNA length and regulates isoform expression." Cell **150**(1): 53-64.

Berzoti-Coelho, M. G., A. F. Ferreira, N. de Souza Nunes, M. T. Pinto, M. C. Junior, B. P. Simoes, A. C. Martinez, E. X. Souto, R. A. Panepucci, D. T. Covas, S. Kashima and F. A. Castro (2016). "The expression of Death Inducer-Obliterator (DIDO) variants in Myeloproliferative Neoplasms." Blood Cells Mol Dis **59**: 25-30.

Beyer, A. L. and Y. N. Osheim (1988). "Splice site selection, rate of splicing, and alternative splicing on nascent transcripts." Genes & Development **2**: 754-765.

Bhatt, D. M., A. Pandya-Jones, A. J. Tong, I. Barozzi, M. M. Lissner, G. Natoli, D. L. Black and S. T. Smale (2012). "Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions." Cell **150**(2): 279-290.

Bieberstein, N. I., F. Carrillo Oesterreich, K. Straube and K. M. Neugebauer (2012). "First exon length controls active chromatin signatures and transcription." Cell Rep **2**(1): 62-68.

Bintu, L., T. Ishibashi, M. Dangkulwanich, Y. Y. Wu, L. Lubkowska, M. Kashlev and C. Bustamante (2012). "Nucleosomal elements that control the topography of the barrier to transcription." Cell **151**(4): 738-749.

Boireau, S., P. Maiuri, E. Basyuk, M. de la Mata, A. Knezevich, B. Pradet-Balade, V. Backer, A. Kornblihtt, A. Marcello and E. Bertrand (2007). "The transcriptional cycle of HIV-1 in real-time and live cells." J Cell Biol **179**(2): 291-304.

Bondarenko, V. A., L. M. Steele, A. Ujvari, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse and V. M. Studitsky (2006). "Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II." Mol Cell **24**(3): 469-479.

Boutz, P. L., A. Bhutkar and P. A. Sharp (2015). "Detained introns are a novel, widespread class of post-transcriptionally spliced introns." Genes Dev **29**(1): 63-80.

Busch, A. and K. J. Hertel (2012). "Evolution of SR protein and hnRNP splicing regulatory factors." Wiley Interdiscip Rev RNA **3**(1): 1-12.

Buxade, M., N. Morrice, D. L. Krebs and C. G. Proud (2008). "The PSF.p54nrb complex is a novel Mnk substrate that binds the mRNA for tumor necrosis factor alpha." J Biol Chem **283**(1): 57-65.

Carmo-Fonseca, M., R. Pepperkok, B. S. Sproat, W. Ansorge, M. S. Swanson and A. I. Lamond (1991). "In vivo detection of snRNP-rich organelles in the nuclei of mammalian cells." EMBO J **10**(7): 1863-1873.

Carrillo Oesterreich, F., S. Preibisch and K. M. Neugebauer (2010). "Global analysis of nascent RNA reveals transcriptional pausing in terminal exons." Mol Cell **40**(4): 571-581.

Castello, A., B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D. T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld and M. W. Hentze (2012). "Insights into RNA biology from an atlas of mammalian mRNA-binding proteins." Cell **149**(6): 1393-1406.

Castello, A., B. Fischer, C. K. Frese, R. Horos, A. M. Alleaume, S. Foehr, T. Curk, J. Krijgsveld and M. W. Hentze (2016). "Comprehensive Identification of RNA-Binding Domains in Human Cells." Mol Cell **63**(4): 696-710.

Castello, A., R. Horos, C. Strein, B. Fischer, K. Eichelbaum, L. M. Steinmetz, J. Krijgsveld and M. W. Hentze (2013). "System-wide identification of RNA-binding proteins by interactome capture." Nat Protoc **8**(3): 491-500.

Cavaloc, Y., C. F. Bourgeois, L. Kister and J. Stévenin (1999). "The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers." RNA **5**: 468–483.

Chen, M. and J. L. Manley (2009). "Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches." Nat Rev Mol Cell Biol **10**(11): 741-754.

Cheng, Y., C. Luo, W. Wu, Z. Xie, X. Fu and Y. Feng (2016). "Liver-Specific Deletion of SRSF2 Caused Acute Liver Failure and Early Death in Mice." Mol Cell Biol **36**(11): 1628-1638.

Cheung, A. C. M. and P. Cramer (2011). "Structural basis of RNA polymerase II backtracking, arrest and reactivation." Nature **471**: 249-253.

Chow LT, G. RE, B. TR and R. RJ. (1977). "An amazing sequence arrangement at the 5'ends of adenovirus 3 messenger RNA." Cell **12**: 1-8.

Corden, J. L. (1990). "Tails of RNA polymerase II." Trends Biochem Sci **15**(10): 383-387.

Custodio, N. and M. Carmo-Fonseca (2016). "Co-transcriptional splicing and the CTD code." Crit Rev Biochem Mol Biol **51**(5): 395-411.

Cvitkovic, I. and M. S. Jurica (2013). "Spliceosome database: a tool for tracking components of the spliceosome." Nucleic Acids Res **41**(Database issue): D132-141.

Daehwan Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg (2013). "METHOD Open Access TopHat2: accurate alignment of Transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biology **14**.

Das, R., J. Yu, Z. Zhang, M. P. Gygi, A. R. Krainer, S. P. Gygi and R. Reed (2007). "SR proteins function in coupling RNAP II transcription to pre-mRNA splicing." Mol Cell **26**(6): 867-881.

De Conti, L., M. Baralle and E. Buratti (2013). "Exon and intron definition in pre-mRNA splicing." Wiley Interdiscip Rev RNA **4**(1): 49-60.

Dujardin, G., C. Lafaille, M. de la Mata, L. E. Marasco, M. J. Munoz, C. Le Jossic-Corcos, L. Corcos and A. R. Kornblihtt (2014). "How slow RNA polymerase II elongation favors alternative exon skipping." Mol Cell **54**(4): 683-690.

Dutertre, M., G. Sanchez, M. C. De Cian, J. Barbier, E. Dardenne, L. Gratadou, G. Dujardin, C. Le Jossic-Corcos, L. Corcos and D. Auboeuf (2010). "Cotranscriptional exon skipping in the genotoxic stress response." Nat Struct Mol Biol **17**(11): 1358-1366.

Dvinge, H., E. Kim, O. Abdel-Wahab and R. K. Bradley (2016). "RNA splicing factors as oncoproteins and tumour suppressors." Nat Rev Cancer **16**(7): 413-430.

Emig, D., N. Salomonis, J. Baumbach, T. Lengauer, B. R. Conklin and M. Albrecht (2010). "AltAnalyze and DomainGraph: analyzing and visualizing exon expression data." Nucleic Acids Res **38**(Web Server issue): W755-762.

Emili, A., M. Shales, S. Mccracken, W. Xie, P. W. Tucker, R. Kobayashi, B. J. Blencowe and C. J. Ingles (2002). "Splicing and transcription-associated proteins PSF and p54nrb/NonO bind to the RNA polymerase II CTD." RNA **8**: 1102-1111.

Fischer, U., A. K. Struss, D. Hemmer, A. Michel, W. Henn, W. I. Steudel and E. Meese (2001). "PHF3 expression is frequently reduced in glioma." Cytogenet Cell Genet **94**(3-4): 131-136.

Flavell, S. W., T. K. Kim, J. M. Gray, D. A. Harmin, M. Hemberg, E. J. Hong, E. Markenscoff-Papadimitriou, D. M. Bear and M. E. Greenberg (2008). "Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection." Neuron **60**(6): 1022-1038.

Fox-Walsh, K. L., Y. Dou, B. J. Lam, S. P. Hung, P. F. Baldi and K. J. Hertel (2005). "The architecture of pre-mRNAs affects mechanisms of splice-site pairing." Proc Natl Acad Sci U S A **102**(45): 16176-16181.

Futterer, A., M. R. Campanero, E. Leonardo, L. M. Criado, J. M. Flores, J. M. Hernandez, J. F. San Miguel and A. C. Martinez (2005). "Dido gene expression alterations are implicated in the induction of hematological myeloid neoplasms." J Clin Invest **115**(9): 2351-2362.

Futterer, A., J. de Celis, R. Navajas, L. Almonacid, J. Gutierrez, A. Talavera-Gutierrez, C. Pacios-Bras, I. Bernascone, F. Martin-Belmonte and A. C. Martinez (2017). "DIDO as a Switchboard that Regulates Self-Renewal and Differentiation in Embryonic Stem Cells." Stem Cell Reports **8**(4): 1062-1075.

Futterer, A., A. Raya, M. Llorente, J. C. Izpisua-Belmonte, J. L. de la Pompa, P. Klatt and A. C. Martinez (2012). "Ablation of Dido3 compromises lineage commitment of stem cells in vitro and during early embryonic development." Cell Death Differ **19**(1): 132-143.

Garcia-Domingo, D., E. Leonardo, A. Grandien, P. Martinez, J. P. Albar, J. C. Izpisua-Belmonte and A. C. Martinez (1999). "DIO-1 is a gene involved in onset of apoptosis in vitro, whose misexpression disrupts limb development." Proc Natl Acad Sci U S A **96**(14): 7992-7997.

Garcia-Domingo, D., D. Ramirez, G. Gonzalez de Buitrago and C. Martinez-A (2003). "Death Inducer-Obliterator 1 Triggers Apoptosis after Nuclear Translocation and Caspase Upregulation." Molecular and Cellular Biology **23**(9): 3216-3225.

Gatchalian, J., A. Futterer, S. B. Rothbart, Q. Tong, H. Rincon-Arano, A. Sanchez de Diego, M. Groudine, B. D. Strahl, A. C. Martinez, K. H. van Wely and T. G. Kutateladze (2013). "Dido3 PHD modulates cell differentiation and division." Cell Rep **4**(1): 148-158.

Gatchalian, J., C. Mora Gallardo, S. A. Shinsky, R. R. Ospina, A. M. Liendo, K. Krajewski, B. J. Klein, F. H. Andrews, B. D. Strahl, M. v. W. KH and T. G. Kutateladze (2016). "Chromatin condensation and recruitment of PHD finger proteins to histone H3K4me3 are mutually exclusive." Nucleic Acids Res **44**(13): 6102-6112.

Giulietti, M., F. Piva, M. D'Antonio, P. D'Onorio De Meo, D. Paoletti, T. Castrignano, A. M. D'Erchia, E. Picardi, F. Zambelli, G. Principato, G. Pavesi and G. Pesole (2013). "SpliceAid-F: a database of human splicing factors and their RNA-binding sites." Nucleic Acids Res **41**(Database issue): D125-131.

Goren, A., O. Ram, M. Amit, H. Keren, G. Lev-Maor, I. Vig, T. Pupko and G. Ast (2006). "Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers." Mol Cell **22**(6): 769-781.

Gornemann, J., K. M. Kotovic, K. Hujer and K. M. Neugebauer (2005). "Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex." Mol Cell **19**(1): 53-63.

Gozani, O., J. G. Patton and R. Reed (1994). "A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction." The EMBO Journal **13**(14): 3356-3367.

Gromak, N., S. West and N. J. Proudfoot (2006). "Pause sites promote transcriptional termination of mammalian RNA polymerase II." Mol Cell Biol **26**(10): 3986-3996.

Guerrero, A. A., M. C. Gamero, V. Trachana, A. Futterer, C. Pacios-Bras, N. P. Diaz-Concha, J. C. Cigudosa, A. C. Martinez and K. H. van Wely (2010). "Centromere-localized breaks indicate the generation of DNA damage by the mitotic spindle." Proc Natl Acad Sci U S A **107**(9): 4159-4164.

Gunderson, S. I., M. Polycarpou-Schwarz and I. W. Mattaj (1998). "U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase." Mol Cell **1**(2): 255-264.

Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A.-C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan and T. Tuschl (2010). "Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP." Cell **141**(1): 129-141.

Han, B., H. K. Park, T. Ching, J. Panneerselvam, H. Wang, Y. Shen, J. Zhang, L. Li, R. Che, L. Garmire and P. Fei (2017). "Human DBR1 modulates the recycling of snRNPs to affect alternative RNA splicing and contributes to the suppression of cancer development." Oncogene **36**(38): 5382-5391.

Hausner, T. P. e. a. (1990). "Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles."

Haynes, C. and L. M. Iakoucheva (2006). "Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins." Nucleic Acids Res **34**(1): 305-312.

Henriques, T., D. A. Gilchrist, S. Nechaev, M. Bern, G. W. Muse, A. Burkholder, D. C. Fargo and K. Adelman (2013). "Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals." Mol Cell **52**(4): 517-528.

Herzel, L., D. S. M. Ottoz, T. Alpert and K. M. Neugebauer (2017). "Splicing and transcription touch base: co-transcriptional spliceosome assembly and function." Nat Rev Mol Cell Biol **18**(10): 637-650.

Hirose, Y., R. Tacke and J. L. Manley (1999). "Phosphorylated RNA polymerase II stimulates pre-mRNA splicing." Genes Dev **13**(10): 1234-1239.

House, A. E. and K. W. Lynch (2008). "Regulation of alternative splicing: more than just the ABCs." J Biol Chem **283**(3): 1217-1221.

Howe, K. J. (2003). "Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae." Rna **9**(8): 993-1006.

Huang, Y., W. Li, X. Yao, Q. J. Lin, J. W. Yin, Y. Liang, M. Heiner, B. Tian, J. Hui and G. Wang (2012). "Mediator complex regulates alternative mRNA processing via the MED23 subunit." Mol Cell **45**(4): 459-469.

Huranová, M., I. Ivani, A. Benda, I. Poser, Y. Brody, M. Hof, Y. Shav-Tal, K. M. Neugebauer and D. Staněk (2010). "The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells." The Journal of Cell Biology **191**(1): 75-86.

Hyslop, S. J., A. M. Duncan, S. Pitkanen and B. H. Robinson (1996). "Assignment of the PSST subunit gene of human mitochondrial complex I to chromosome 19p13." Genomics **37**(3): 375-380.

Iannone, C. and J. Valcarcel (2013). "Chromatin's thread to alternative splicing regulation." Chromosoma **122**(6): 465-474.

Iniguez, L. P. and G. Hernandez (2017). "The Evolutionary Relationship between Alternative Splicing and Gene Duplication." Front Genet **8**: 14.

Inoue, D., M. Stemmer, T. Thumberger, T. Ruppert, F. Barenz, J. Wittbrodt and O. J. Gruss (2017). "Expression of the novel maternal centrosome assembly factor Wdr8 is required for vertebrate embryonic mitoses." Nat Commun **8**: 14090.

Ip, J. Y., D. Schmidt, Q. Pan, A. K. Ramani, A. G. Fraser, D. T. Odom and B. J. Blencowe (2011). "Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation." Genome Res **21**(3): 390-401.

Jarvelin, A. I., M. Noerenberg, I. Davis and A. Castello (2016). "The new (dis)order in RNA regulation." Cell Commun Signal **14**: 9.

Jelen, N., J. Ule, M. Zivin and R. B. Darnell (2007). "Evolution of Nova-dependent splicing regulation in the brain." PLoS Genet **3**(10): 1838-1847.

Johnson, M. L., A. A. Nagengast and H. K. Salz (2010). "PPS, a large multidomain protein, functions with sex-lethal to regulate alternative splicing in Drosophila." PLoS Genet **6**(3): e1000872.

Kaida, D., M. G. Berg, I. Younis, M. Kasim, L. N. Singh, L. Wan and G. Dreyfuss (2010). "U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation." Nature **468**(7324): 664-668.

Kasahara, M. (2007). "The 2R hypothesis: an update." Curr Opin Immunol **19**(5): 547-552.

Keren, H., G. Lev-Maor and G. Ast (2010). "Alternative splicing and evolution: diversification, exon definition and function." Nat Rev Genet **11**(5): 345-355.

Kim, E., A. Goren and G. Ast (2008). "Alternative splicing: current perspectives." Bioessays **30**(1): 38-47.

Kinkelin, K., G. G. Wozniak, S. B. Rothbart, M. Lidschreiber, B. D. Strahl and P. Cramer (2013). "Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue." Proc Natl Acad Sci U S A **110**(38): 15277-15282.

Knott, G. J., C. S. Bond and A. H. Fox (2016). "The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold." Nucleic Acids Res **44**(9): 3989-4004.

Koga, M., M. Hayashi and D. Kaida (2015). "Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD." Nucleic Acids Res **43**(17): 8258-8267.

Konarska, M. M. a. S., P.A. (1988). "Association of U2, U4, U5, and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA.".

Kornblihtt, A. R., I. E. Schor, M. Allo, G. Dujardin, E. Petrillo and M. J. Munoz (2013). "Alternative splicing: a pivotal step between eukaryotic transcription and translation." Nat Rev Mol Cell Biol **14**(3): 153-165.

Lacadie, S. A. and M. Rosbash (2005). "Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast." Mol Cell **19**(1): 65-75.

Lee, J. A., Z. Z. Tang and D. L. Black (2009). "An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons." Genes Dev **23**(19): 2284-2293.

Lee, J. H. and D. G. Skalnik (2012). "Rbm15-Mkl1 interacts with the Setd1b histone H3-Lys4 methyltransferase via a SPOC domain that is required for cytokine-independent proliferation." PLoS One **7**(8): e42965.

Lee, M., A. Sadowska, I. Bekere, D. Ho, B. S. Gully, Y. Lu, K. S. Iyer, J. Trewhella, A. H. Fox and C. S. Bond (2015). "The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation." Nucleic Acids Res **43**(7): 3826-3840.

Li, S., Z. Li, F. J. Shu, H. Xiong, A. C. Phillips and W. S. Dynan (2014). "Double-strand break repair deficiency in NONO knockout murine embryonic fibroblasts and compensation by spontaneous upregulation of the PSPC1 paralog." Nucleic Acids Res **42**(15): 9771-9780.

Licatalosi, D. D., G. Geiger, M. Minet, S. Schroeder, K. Cilli, J. B. McNeil and D. L. Bentley (2002). "Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II." Mol Cell **9**(5): 1101-1111.

Liu, S. and C. Cheng (2013). "Alternative RNA splicing and cancer." Wiley Interdiscip Rev RNA **4**(5): 547-566.

Luco, R. F., Q. Pan, K. Tominaga, B. J. Blencowe, O. M. Pereira-Smith and T. Misteli (2010). "Regulation of alternative splicing by histone modifications." Science **327**(5968): 996-1000.

Lunde, B. M., C. Moore and G. Varani (2007). "RNA-binding proteins: modular design for efficient function." Nat Rev Mol Cell Biol **8**(6): 479-490.

Martinez, A. C. and K. H. van Wely (2011). "Centromere fission, not telomere erosion, triggers chromosomal instability in human carcinomas." Carcinogenesis **32**(6): 796-803.

Mayr, C. and D. P. Bartel (2009). "Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells." Cell **138**(4): 673-684.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. D.Patterson, M. Wickens and D. L.Bentley (1997). "The C-terminal domain of RNA polymerase II couples mRNA processing to transcription." Nature **385**.

Medenbach, J., S. Schreiner, S. Liu, R. Luhrmann and A. Bindereif (2004). "Human U4/U6 snRNP recycling factor p110: mutational analysis reveals the function of the tetratricopeptide repeat domain in recycling." Mol Cell Biol **24**(17): 7392-7401.

Montes, M., S. Becerra, M. Sanchez-Alvarez and C. Sune (2012). "Functional coupling of transcription and splicing." Gene **501**(2): 104-117.

Natalizio, B. J., N. D. Robson-Dixon and M. A. Garcia-Blanco (2009). "The Carboxyl-terminal Domain of RNA Polymerase II Is Not Sufficient to Enhance the Efficiency of Pre-mRNA Capping or Splicing in the Context of a Different Polymerase." J Biol Chem **284**(13): 8692-8702.

Neugebauer, K. M. (2002). "On the importance of being co-transcriptional." Journal of Cell Science **115**(20): 3865-3871.

Niibori, Y., F. Hayashi, K. Hirai, M. Matsui and K. Inokuchi (2007). "Alternative poly(A) site-selection regulates the production of alternatively spliced vesl-1/homer1 isoforms that encode postsynaptic scaffolding proteins." Neurosci Res **57**(3): 399-410.

Nogues, G., S. Kadener, P. Cramer, D. Bentley and A. R. Kornblihtt (2002). "Transcriptional activators differ in their abilities to control alternative splicing." J Biol Chem **277**(45): 43110-43114.

Nojima, T., T. Gomes, M. Carmo-Fonseca and N. J. Proudfoot (2016). "Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide." Nat Protoc **11**(3): 413-428.

Nudler, E. (2012). "RNA polymerase backtracking in gene regulation and genome instability." Cell **149**(7): 1438-1445.

O'Brien, K., A. J. Matlin, A. M. Lowell and M. J. Moore (2008). "The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing." J Biol Chem **283**(48): 33147-33154.

Oesterreich, F. C., L. Herzel, K. Straube, K. Hujer, J. Howard and K. M. Neugebauer (2016). "Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II." Cell **165**(2): 372-381.

Oltean, S. and D. O. Bates (2014). "Hallmarks of alternative splicing in cancer." Oncogene **33**(46): 5311-5318.

Palangat, M., C. T. Hittinger and R. Landick (2004). "Downstream DNA selectively affects a paused conformation of human RNA polymerase II." J Mol Biol **341**(2): 429-442.

Pandit, S., Y. Zhou, L. Shiue, G. Coutinho-Mansfield, H. Li, J. Qiu, J. Huang, G. W. Yeo, M. Ares, Jr. and X. D. Fu (2013). "Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing." Mol Cell **50**(2): 223-235.

Patton, J. G., E. B. Porro, J. Galceran, P. Tempst and B. Nadal-Ginard (1993). "Cloning and characterization of PSF, a novel pre-mRNA splicing factor." Genes Dev **7**(3): 393-406.

Peng, R., B. T. Dye, I. Pérez, D. C. Barnard, A. B. Thompson and J. G. Patton (2002). "PSF and p54nrb bind a conserved stem in U5 snRNA." RNA **8**: 1334-1347.

Peng, R., I. Hawkins, A. J. Link and J. G. Patton (2014). "The Splicing Factor PSF is Part of a Large Complex That Assembles in the Absence of pre-mRNA and Contains All 5 snRNPs." RNA Biology **3**(2): 69-76.

Pinskaya, M., Y. Ghavi-Helm, S. Mariotte-Labarre, A. Morillon, J. Soutourina and M. Werner (2014). "PHD and TFIIS-Like domains of the Bye1 transcription factor determine its multivalent genomic distribution." PLoS One **9**(7): e102464.

Pradeepa, M. M., H. G. Sutherland, J. Ule, G. R. Grimes and W. A. Bickmore (2012). "Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing." PLoS Genet **8**(5): e1002717.

Prieto, I., A. Kouznetsova, A. Futterer, V. Trachana, E. Leonardo, A. Alonso Guerrero, M. Cano Gamero, C. Pacios-Bras, H. Leh, M. Buckle, M. Garcia-Gallo, L. Kremer, A. Serrano, F. Roncal, J. P. Albar, J. L. Barbero, A. C. Martinez and K. H. van Wely (2009). "Synaptonemal complex assembly and H3K4Me3 demethylation determine DIDO3 localization in meiosis." Chromosoma **118**(5): 617-632.

Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." Nat Protoc **8**(11): 2281-2308.

Rino, J. and M. Carmo-Fonseca (2009). "The spliceosome: a self-organized macromolecular machine in the nucleus?" Trends Cell Biol **19**(8): 375-384.

Roberts, G. C., C. Gooding, H. Y. Mak, N. J. Proudfoot and C. W. Smith (1998). "Co-transcriptional commitment to alternative splice site selection." Nucleic Acids Res **26**(24): 5568-5572.

Rojas, A. M., L. Sanchez-Pulido, A. Futterer, K. H. van Wely, A. C. Martinez and A. Valencia (2005). "Death inducer obliterator protein 1 in the context of DNA regulation. Sequence analyses of distant homologues point to a novel functional role." FEBS J **272**(14): 3505-3511.

Romanelli, M. G., E. Diani and P. M. Lievens (2013). "New insights into functional roles of the polypyrimidine tract-binding protein." Int J Mol Sci **14**(11): 22906-22932.

Rosonina, E., J. Y. Ip, J. A. Calarco, M. A. Bakowski, A. Emili, S. McCracken, P. Tucker, C. J. Ingles and B. J. Blencowe (2005). "Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo." Mol Cell Biol **25**(15): 6734-6746.

Ruskin, B., P. D. Zamore and M. R. Green (1988). "A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly." Cell **52**(2): 207-219.

Ruthenburg, A. J., C. D. Allis and J. Wysocka (2007). "Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark." Mol Cell **25**(1): 15-30.

Sahoo, M. and S. Klumpp (2013). "Backtracking dynamics of RNA polymerase: pausing and error correction." J Phys Condens Matter **25**(37): 374104.

Saldi, T., M. A. Cortazar, R. M. Sheridan and D. L. Bentley (2016). "Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing." J Mol Biol **428**(12): 2623-2635.

Sanchez de Diego, A., A. Alonso Guerrero, A. C. Martinez and K. H. van Wely (2014). "Dido3-dependent HDAC6 targeting controls cilium size." Nat Commun **5**: 3500.

Sanchez-Pulido, L., A. M. Rojas, K. H. van Wely, A. C. Martinez and A. Valencia (2004). "SPOC: a widely distributed domain associated with cancer, apoptosis and transcription." BMC Bioinformatics **5**: 91.

Sandberg, R., J. R. Neilson, A. Sarma, P. A. Sharp and C. B. Burge (2008). "Proliferating Cells Express mRNAs with Shortened 3′ Untranslated Regions and Fewer MicroRNA Target Sites." Science **320**: 1643-1647.

Schaal, T. D. and T. Maniatis (1999). "Selection and Characterization of Pre-mRNA Splicing Enhancers: Identification of Novel SR Protein-Specific Enhancer Sequences." Molecular and Cellular Biology **19**(3): 1705–1719.

Schor, I. E., N. Rascovan, F. Pelisch, M. Allo and A. R. Kornblihtt (2009). "Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing." Proc Natl Acad Sci U S A **106**(11): 4325-4330.

Schroeder, S. C., B. Schwer, S. Shuman and D. Bentley (2000). "Dynamic association of capping enzymes with transcribing RNA polymerase II." Genes Dev **14**(19): 2435-2440.

Schwartz, S., E. Meshorer and G. Ast (2009). "Chromatin organization marks exon-intron structure." Nat Struct Mol Biol **16**(9): 990-995.

Schwer, B. and S. Shuman (2011). "Deciphering the RNA polymerase II CTD code in fission yeast." Mol Cell **43**(2): 311-318.

Shandilya, J. and S. G. Roberts (2012). "The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling." Biochim Biophys Acta **1819**(5): 391-400.

Shav-Tal, Y., J. Blechman, X. Darzacq, C. Montagna, B. T. Dye, J. G. Patton, R. H. Singer and D. Zipori (2005). "Dynamic sorting of nuclear components into distinct nucleolar caps during transcriptional inhibition." Molecular Biology of the cell **16**(5): 2395-2413.

Sims, R. J., 3rd, R. Belotserkovskaya and D. Reinberg (2004). "Elongation by RNA polymerase II: the short and long of it." Genes Dev **18**(20): 2437-2468.

Sims, R. J., 3rd, S. Millhouse, C. F. Chen, B. A. Lewis, H. Erdjument-Bromage, P. Tempst, J. L. Manley and D. Reinberg (2007). "Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing." Mol Cell **28**(4): 665-676.

Singh, B. and E. Eyras (2017). "The role of alternative splicing in cancer." Transcription **8**(2): 91-98.

Sobell, H. M. (1985). "Actinomycin and DNA transcription." Proc Natl Acad Sci U S A **82**(16): 5328-5331.

Soille, P. and L. M. Vincent (1990). Determining watersheds in digital pictures via flooding simulations, SPIE.

Spies, N., C. B. Nielsen, R. A. Padgett and C. B. Burge (2009). "Biased chromatin signatures around polyadenylation sites and exons." Mol Cell **36**(2): 245-254.

Spiluttini, B., B. Gu, P. Belagal, A. S. Smirnova, V. T. Nguyen, C. Hebert, U. Schmidt, E. Bertrand, X. Darzacq and O. Bensaude (2010). "Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells." J Cell Sci **123**(Pt 12): 2085-2093.

Srivastava, C., K. Irshad, B. Dikshit, P. Chattopadhyay, C. Sarkar, D. K. Gupta, S. Sinha and K. Chosdol (2018). "FAT1 modulates EMT and stemness genes expression in hypoxic glioblastoma." International Journal of Cancer **142**(4): 805-812.

Struss, A.-K., B. Romeike, A. Munnia, W. Nastainczyk, W.-I. Steudel, J. König, H. Ohgaki, W. Feiden, U. Fischer and E. Meese (2001). "PHF3-specific antibody responses in over 60% of patients with glioblastoma multiforme." Oncogene **20**: 4107-4114.

Tacke, R. and J. L.Manley (1995). "The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities." The EMBO Journal **14**(14): 3540-3551.

Takeuchi, A., K. Iida, T. Tsubota, M. Hosokawa, M. Denawa, J. B. Brown, K. Ninomiya, M. Ito, H. Kimura, T. Abe, H. Kiyonari, K. Ohno and M. Hagiwara (2018). "Loss of Sfpq Causes Long-Gene Transcriptopathy in the Brain." Cell Rep **23**(5): 1326-1341.

Tardiff, D. F., S. A. Lacadie and M. Rosbash (2006). "A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional." Mol Cell **24**(6): 917-929.

Thomas-Jinu, S., P. M. Gordon, T. Fielding, R. Taylor, B. N. Smith, V. Snowden, E. Blanc, C. Vance, S. Topp, C. H. Wong, H. Bielen, K. L. Williams, E. P. McCann, G. A. Nicholson, A. Pan-Vazquez, A. H. Fox, C. S. Bond, W. S. Talbot, I. P. Blair, C. E. Shaw and C. Houart (2017). "Non-nuclear Pool of Splicing Factor SFPQ Regulates Axonal Transcripts Required for Normal Motor Development." Neuron **94**(2): 322-336 e325.

Tian, H. and R. Kole (2001). "Strong RNA splicing enhancers identified by a modified method of cycled selection interact with SR protein." J Biol Chem **276**(36): 33833-33839.

Tilgner, H., C. Nikolaou, S. Althammer, M. Sammeth, M. Beato, J. Valcarcel and R. Guigo (2009). "Nucleosome positioning as a determinant of exon recognition." Nat Struct Mol Biol **16**(9): 996-1001.

Trachana, V., K. H. van Wely, A. A. Guerrero, A. Futterer and A. C. Martinez (2007). "Dido disruption leads to centrosome amplification and mitotic checkpoint defects compromising chromosome stability." Proc Natl Acad Sci U S A **104**(8): 2691-2696.

Tsui, S., T. Dai, S. Roettger, W. Schempp, E. C. Salido and P. H. Yen (2000). "Identification of two novel proteins that interact with germ-cell-specific RNA-binding proteins DAZ and DAZL1." Genomics **65**(3): 266-273.

Tyagi, A., J. Ryme, D. Brodin, A. K. Ostlund Farrants and N. Visa (2009). "SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing." PLoS Genet **5**(5): e1000470.

Ule, J., K. B. Jensen, M. Ruggiu, A. Mele, A. Ule and R. B. Darnell (2003). "CLIP identifies Nova-regulated RNA networks in the brain." Science **302**(5648): 1212-1215.

van Wely, K. H., C. Mora Gallardo, K. R. Vann and T. G. Kutateladze (2017). "Epigenetic countermarks in mitotic chromosome condensation." Nucleus **8**(2): 144-149.

Villares, R., J. Gutierrez, A. Futterer, V. Trachana, F. Gutierrez del Burgo and A. C. Martinez (2015). "Dido mutations trigger perinatal death and generate brain abnormalities and behavioral alterations in surviving adult mice." Proc Natl Acad Sci U S A **112**(15): 4803-4808.

Wahl, M. C. and R. Lührmann (2015). "SnapShot: Spliceosome Dynamics I." Cell **161**(6): 1474-1474.e1471.

Wahl, M. C., C. L. Will and R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-718.

Wan, L. and G. Dreyfuss (2017). "Splicing-Correcting Therapy for SMA." Cell **170**(1): 5.

Wang, E. T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth and C. B. Burge (2008). "Alternative isoform regulation in human tissue transcriptomes." Nature **456**(7221): 470-476.

Wassarman, D. A. a. S., J.A. (1992). "Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing."

Weibrecht, I., K. J. Leuchowius, C. M. Clausson, T. Conze, M. Jarvius, W. M. Howell, M. Kamali-Moghaddam and O. Soderberg (2010). "Proximity ligation assays: a recent addition to the proteomics toolbox." Expert Rev Proteomics **7**(3): 401-409.

Will, C. L. and R. Luhrmann (2011). "Spliceosome structure and function." Cold Spring Harb Perspect Biol **3**(7).

Yao, C., J. Biesinger, J. Wan, L. Weng, Y. Xing, X. Xie and Y. Shi (2012). "Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation." Proc Natl Acad Sci U S A **109**(46): 18773–18778.

Zevini, A., D. Olagnier and J. Hiscott (2017). "Crosstalk between Cytoplasmic RIG-I and STING Sensing Pathways." Trends in Immunology **38**(3): 194-205.

Zhang, J. and J. L. Manley (2013). "Misregulation of pre-mRNA alternative splicing in cancer." Cancer Discov **3**(11): 1228-1237.

# ANNEXES

## Publications:

**1.** Gatchalian, J.*, **C. Mora Gallardo***, S. A. Shinsky, R. R. Ospina, A. M. Liendo, K. Krajewski, B. J. Klein, F. H. Andrews, B. D. Strahl, K. H. van Wely and T. G. Kutateladze (2016). "Chromatin condensation and recruitment of PHD finger proteins to histone H3K4me3 are mutually exclusive." Nucleic Acids Res **44**(13): 6102-6112. *These authors contributed equally to this work.

**2.** K. H. van Wely, **C. Mora Gallardo**, K. R. Vann and T. G. Kutateladze (2017). "Epigenetic countermarks in mitotic chromosome condensation." Nucleus **8**(2): 144-149.

**3. C. Mora Gallardo**, Sánchez de Diego A., J. Gutierrez, A. Talavera-Gutierrez, Fischer T., A. C. Martinez and K. H. van Wely (2019). "Dido3-dependent SFPQ recruitment maintains efficiency in mammalian alternative splicing" Nucleic Acids Res (*Under review*)